

NASA LANGLEY RESEARCH CENTER

GRANT NUMBER: NAG-1-03022

**SITUATION AWARENESS IMPLICATIONS OF ADAPTIVE
AUTOMATION OF AIR TRAFFIC CONTROLLER
INFORMATION PROCESSING FUNCTIONS**

FINAL REPORT FOR NORTH CAROLINA STATE UNIVERSITY

Prepared By:

David B. Kaber, Christopher K. McClernon, Carlene M. Perry and Noa Segall

Period of Performance: December 1, 2002 to May 15, 2004

INTRODUCTORY INFORMATION

Legal Name of Organization: North Carolina State University

Address: Sponsored Programs Services
North Carolina State University
P.O. 7514
Raleigh, NC 27695-7514

Originating Unit: Department of Industrial Engineering

Organization Type: Educational

Principal Investigator:
Name: David B. Kaber
Phone: (919) 515-3086

Project Title: A Situation Awareness-Based Approach to Adaptive Automation

Technical Contact: Lawrence J. Prinzel, NASA Airspace Systems Program/Aviation Safety Program

Starting Date: 12/1/02

Expiration Date: 5/15/04

Elapsed Project Duration: 17.5 mos.

A Situation Awareness-Based Approach to Adaptive Automation

Executive Summary

The goal of this research was to define a measure of situation awareness (SA) in an air traffic control (ATC) task and to assess the influence of adaptive automation (AA) of various information processing functions on controller perception, comprehension and projection. The measure was also to serve as a basis for defining and developing an approach to triggering dynamic control allocations, as part of AA, based on controller SA.

To achieve these objectives, an enhanced version of an ATC simulation (Multitask©) was developed for use in two human factors experiments. The simulation captured the basic functions of Terminal Radar Approach Control (TRACON) and was capable of presenting to operators four different modes of control, including information acquisition, information analysis, decision making and action implementation automation, as well as a completely manual control mode. The SA measure that was developed as part of the research was based on the Situation Awareness Global Assessment Technique (SAGAT), previous goal-directed task analyses of enroute control and TRACON, and a separate cognitive task analysis on the ATC simulation. The results of the analysis on Multitask were used as a basis for formulating SA queries as part of the SAGAT-based approach to measuring controller SA, which was used in the experiments.

A total of 16 subjects were recruited for both experiments. Half the subjects were used in Experiment #1, which focused on assessing the sensitivity and reliability of the SA measurement approach in the ATC simulation. Comparisons were made of manual versus automated control. The remaining subjects were used in the second experiment, which was intended to more completely describe the SA implications of AA applied to specific controller information processing functions, and to describe how the measure could ultimately serve as a trigger of dynamic function allocations in the application of AA to ATC. Comparisons were made of the sensitivity of the SA measure to automation manipulations impacting both higher-order information processing functions, such as information analysis and decision making, versus lower-order functions, including information acquisition and action implementation. All subjects were exposed to all forms of AA of the ATC task and the manual control condition. The approach to AA used in both experiments was to match operator workload, assessed using a secondary task, to dynamic control allocations in the primary task. In total, the subjects in each experiment participated in 10 trials with each lasting between 45 minutes and 1 hour. In both experiments, ATC performance was measured in terms of aircraft cleared, conflicting, and collided. Secondary task (gauge monitoring) performance was assessed in terms of a hit-to-signal ratio. As part of the SA measure, three simulation freezes were conducted during each trial to administer queries on Level 1, 2, and 3 SA.

Results revealed ATC performance to be significantly superior when automation was applied to lower-order sensory processing functions, including information acquisition and action implementation, as compared to higher-order functions, specifically information analysis. There were also significant differences among manual control periods as part of AA of the various information processing functions and across both studies. When AA was applied to the information analysis function, performance during the manual control periods was superior to manual performance as part of AA of action implementation. Only the second experiment revealed significant effects of the various forms of AA on workload, or secondary task performance. Most interestingly, the higher levels of automation, including information analysis and decision making, appeared to cause worse secondary-task performance or higher workload. This may be attributed to the visual demand of the interface displays as part of the automation, as well as operator evaluation of automation recommendations relative to their own task strategy. Finally, the first experiment, which involved a straightforward application of a SAGAT-based approach to measuring SA, did not reveal significant differences among the various AA conditions in terms of operator perception,

comprehension, and projection. However, in the second experiment, subject recall of aircraft was queued using a graphical aid and relevance weights were assigned to aircraft at the time of simulation freezes. These modifications lead to sensitivity of the measurement approach to the AA manipulations. In particular, operator perception and total SA improved with the application of AA to the information acquisition function, providing operators with assistance in identifying potential aircraft conflicts. Finally, we described how the new measurement technique could be used to facilitate a SA-matched approach to AA of ATC.

Table of Contents

1. Introduction	1
Background on Air Traffic Control Automation	2
<i>Advantages and Disadvantages of Automation</i>	3
Adaptive Automation Research in the Context of ATC	5
Research on Measures of SA and AA.....	6
2. Objectives and Overview of Research	8
Enhancement of the Multitask" Simulation	8
Defining a Measure of SA Sensitive to Dynamic Function Allocations as part of AA15	
Quantifying the impact of AA on SA in an ATC simulation	18
3. Experiment #1.....	18
Participants	18
Tasks	19
Experiment Equipment.....	20
Approach to AA.....	20
Experimental Design.....	21
Variables.....	21
Procedures	22
Data Analysis.....	24
<i>Multitask Data</i>	25
<i>Workload Data</i>	25
<i>Diagnostics and Statistical Model</i>	25
<i>Situation Awareness Data</i>	26
Specific Hypotheses.....	27
Results and Discussion.....	28
<i>Primary Task Performance</i>	28
<i>Workload (Secondary-task Performance)</i>	30
<i>Situation Awareness</i>	31
4. Experiment #2.....	33
Participants	35
Tasks and Equipment.....	35
Approach to AA.....	35
Design of Experiment and Variables	36
Procedures	37
Data Analysis.....	37
<i>Diagnostics and statistical model</i>	38
Specific Hypotheses.....	38
Results and Discussion.....	39
<i>Primary Task Performance</i>	39
<i>Workload (Secondary-task Performance)</i>	42
<i>Situation Awareness</i>	45
<i>Performance and Situation Awareness Correlation Analyses</i>	47

5. Conclusions	48
Caveats	48
<i>Simulation Fidelity</i>	48
<i>Subject Training</i>	49
<i>SA Measurement Technique Problems</i>	49
Design Implications for Future ATC Automation	49
<i>Contributions to AA Research</i>	49
<i>Contributions to SA Research</i>	50
Future Research	51
References	55
Appendix A: GDTA for Multitask” simulation.....	58
Appendix B: SAGAT queries.....	65

A Situation Awareness-Based Approach to Adaptive Automation

1. Introduction

Adaptive automation (AA) refers to complex systems in which the level of automation or the number of system functions being automated can be modified in real time (Scerbo, 1996). Recent research (Hilburn, Jorna, Byrne & Parasuraman, 1997; Kaber, Riley, Tan & Endsley, 2001; Parasuraman, 2000) has identified AA as a potential solution to complex system operator out-of-the-loop (OOTL) performance and situation awareness (SA) problems associated with static, high-level automation (e.g., supervisory control systems). It has been suggested that facilitating dynamic allocations of system control functions to a human operator or computer over time can moderate operator workload and, at the same time, facilitate SA, or operator preparedness for unexpected system states, by maintaining some level of operator involvement in control loops.

Unfortunately, little research has been conducted on the SA implications of AA. Two other recent studies attempted to describe SA under AA in dynamic control tasks, including radar monitoring and remote control of roving robots (Kaber & Endsley, 2004; Kaber, Wright & Hughes, 2002). However, both of these experiments explored model-based approaches to AA; that is, the dynamic function allocations (DFAs) during task performance (shifts from human manual operation to some form of automation (e.g., shared control, decision support, etc.)) were programmed based on anticipated changes in operator workload due to scheduled task events. (We review these studies in detail later in the report.) Although model-based approaches to AA, have been demonstrated to be effective for supporting monitoring task performance (Parasuraman, Mouloua, Molloy & Hilburn, 1993; Hilburn, Molloy, Wong & Parasuraman, 1993), pre-programmed system control allocations cannot be considered to represent a truly adaptive system that changes states based on, for example, real-time task information or operator workload measurement. Another drawback of the research conducted by Kaber & Endsley (2004) is that the task they used was an abstract (laboratory) simulation with limited resemblance to real-world military radar monitoring or air traffic control (ATC). Research is needed to evaluate SA when operators are exposed to truly adaptive systems versus systems employing "arbitrary" automation. In addition, the SA implications of AA need to be examined in realistic simulations of complex systems.

Other approaches to AA in complex systems control include monitoring operator workload states and triggering control allocations based on workload fluctuations, or triggering DFAs on the basis of critical events (e.g., system failures). Up to this point in time, no measures of SA have been developed to assess operator perception, comprehension and projection of system states as a basis for triggering, or driving, DFAs in adaptive systems. A number of studies have demonstrated the effectiveness of AA involving DFAs based on real-time measures of operator workload (e.g., Kaber & Riley, 1999). It is possible that a SA-based approach to AA may produce different results in terms of human performance and workload in controlling adaptive systems, as compared to the other strategies that have already been explored. Research has suggested that a SA measure of the impact of AA on complex system operators may provide different information on operator states than performance and workload measures and may be important to consider in assessing the effectiveness of AA (or the suitability of specific types of AA to complex system operations) (Kaber & Endsley, 2004). The need for development of a SA

measure to serve as a basis for driving AA in complex systems control is yet another motivation for preliminary assessments of the implications of AA on SA. It should be determined whether sensitive measures of SA can be defined to consistently identify changes in operator system knowledge and performance due to dynamic allocations of control modes.

Such measures should also be used to describe the effects of AA on operator SA, when applied to various human-machine system information processing functions. These are important needs because highly complex operations like ATC may involve acute changes in operator functional responsibilities during task performance, which may have unique and critical effects on SA in comparison to workload or performance. Studying SA in adaptive system control and the use of a SA-based approach to prescribing DFAs may be important because operator out-of-the-loop unfamiliarity (OOTLUF) can occur in adaptive systems, as in the use of static automation, depending upon the extent of automation (level and duration). Consequently, operator SA may be degraded. Since OOTLUF can have serious implications on performance, including the ability to take control of a system when errors occur, or to recover systems from failure, considering SA as a basis for making automation decisions could be useful. The present work focused on the former needs, including describing the implications of AA on SA in complex systems control and defining a measure of SA in the context of an ATC simulation with sensitivity to DFAs.

Background on Air Traffic Control Automation

Air traffic control requires high levels of cognitive processing, and one approach for alleviating the stress and workload among controllers is to allocate some controller activities to automation (National Resource Council (NRC), 1997; Parasuraman & Riley, 1997). Air traffic control automation first appeared in the 1960's in the form of the Automated Radar Terminal System (ARTS), which incorporated all aircraft information, within one controlled airspace, into one radar display. Modern versions of the ARTS are still in place today.

There are currently many different forms of automation incorporated in ATC. Wickens (1992, p. 531-532) identified three general types of automation, including automation that can be used to perform functions that the human operator cannot perform because of inherent limitations. Examples of this form of automation in ATC include flight data processing (FDP) for en route centers where information about flights and controller interactions is stored in a database for later data synthesis, data presentation, and computations to be used in performance evaluation and training. Another example of this form of automation is smoothing of aircraft flight information on ATC displays. Flight data regarding specific aircraft is gathered incrementally and presented to controllers on displays, but this unprocessed information causes jagged and irregular movements of icons on control displays, which can be distracting to controllers. An automated process called smoothing extrapolates flight data into smooth, fluid movement information to drive aircraft icons on control displays (NRC, 1997).

Wickens (1992) also describes automation used to perform functions that humans perform poorly or at the cost of high workload. For example, automated systems have been implemented in ATC to facilitate "hand-offs" of aircraft between ATC control sectors, thereby, eliminating the need for verbal communication between controllers as an aircraft passes through airspace, and decreasing controller workload. In addition, controllers were previously required to obtain

aircraft altitude, speed, and other information verbally from pilots for “hand-offs”. They were also expected to perform the complex process of determining specific aircraft headings based on limited and inaccurate aircraft information presented on control displays. This information including aircraft call sign, type of aircraft, destination airport, navigational fixes, altitude, and ground speed are all now automatically presented to the controller in an accurate data tag embedded in the radar display (NRC, 1997).

Automation is also used to augment or assist humans in areas in which they show performance limitations (Wickens, 1992). In an effort to aid controllers with automation, the Federal Aviation Administration (FAA) introduced the center-TRACON automation system (CTAS). This system assists controllers in successfully clearing aircraft, rather than implementing fully automated systems (Dillingham, 1998). The CTAS has four components, including the Traffic Management Advisor (TMA). This computer system gathers information about all aircraft inbound for a landing, and uses the aircraft parameters to develop a plan including a sequence and schedule for landing the aircraft. This plan may be accepted or adjusted by the controller to meet specific requirements not considered by the automated system (e.g., emergency procedures, special requests, etc.; NRC, 1998). Secondly, the CTAS system incorporates a Descent Advisor (DA), which uses aircraft type, capabilities, atmospheric conditions, and a plan from the TMA to advise controllers about descent rates, speeds, and durations for each aircraft (Hilburn et al., 1997). The Final Approach Spacing Tool (FAST) provides controllers with aircraft sequence, runway assignments, speed, and heading advisories. The final approach portion of a flight is the most crucial segment given the close proximity of aircraft near an airport, limited time for controllers to make decisions, and the precision and safety requirements of decisions (NRC, 1997). The FAST has the capability to quickly and precisely adjust to dynamic situations during the final approach segment. (The DA and FAST advisories are presented automatically to the controller through the aircraft data tags on the radar display.) The Expedite Departure Path (EDP) program is a recent addition to the CTAS, which incorporates the same capabilities of the FAST and DA, but for directing aircraft that are departing a controlled area. In addition, these capabilities are becoming available, not only for the aircraft arriving and departing the controlled airport, but for aircraft arriving and departing smaller airports within close proximity to the controlled airport (NRC, 1998; Nolan, 1999).

Other ATC automation systems utilized to assist controllers are the Minimum Safe Altitude Warning (MSAW) system, the Conflict Alert (CA) system, and the automated (aircraft) track deviation system (NRC, 1997). Converging runway display aids (CRDA) have also been developed as part of automated ATC systems. The CRDA projects the future paths of two aircraft landing on converging runways, relieving the controller of having to mentally project the aircraft paths. All of these systems may alleviate controller workload by allocating some task functions to computers (NRC, 1997; Nolan, 1999).

Advantages and Disadvantages of Automation

Automation can have a profound effect on human performance. The forms of automation described above have many potential advantages for controllers including a reduced workload (Laois & Giannacourou, 1995), an increased system reliability (NRC, 1998), and an increased capability to perform complex computations and data management (Wickens & Hollands, 2000).

For example, Laois and Giannacourou (1995) conducted research to determine the impact of five currently automated ATC systems, including: Electronic Data Display (EDD), Trajectory Prediction Aids (TPA), Conflict Detection Aids (CDA), Data Links Applications (DL), and Clearance Advisory Aids (CAA). Their study included observations, interviews, and presenting expert controllers with questionnaires pertaining to the automated systems. They determined that all five forms of automation increased formalization (of communication), performance, and flexibility while decreasing workload by assuming responsibility for ATC functions. They also stated that “significant workload reductions will be effected by aiding decision making and predictive activities more than by automation of routine data acquisition and communication activities” (Laois & Giannacourou, 1995, p. 395).

However, automation in ATC can also present many disadvantages (Dillingham, 1998), including a loss of controller SA (Endsley, 1995a; Endsley & Jones, 1995). As machines perform more and more ATC functions and operations, controllers have less interaction with the system and may become less aware of system operations. OOTL performance may reduce a controller’s ability to detect a problem has occurred, determine the current state of the system, understand what has happened and what courses of actions are needed, and react to the situation (Endsley, 1996). Laois and Giannacourou (1995) also observed this in their study of expert controllers when they found that automation, which placed the controller OOTL (even for short periods), resulted in an inability to react to an emergency situation. Maintaining SA in ATC is critical for accurate decision making and performance (Endsley, 1996). Automation of different human-machine system information processing functions may also affect SA in different ways. Endsley and Kiris (1995) and Kaber, Onal, and Endsley (1999) determined that low-level automation (e.g., information acquisition and action implementation) can improve SA, while high-level automation, like decision-making aids, can decrease SA. Examples of ATC automation leading to decrements in controller SA can be found in controller mode errors. Complex automation integrating multiple modes of operation can lead to situations in which a controller is unaware of the mode in which the system is currently operating; thus, the controller inputs data, or extrapolates data from a system, assuming the wrong system mode is active. Many aircraft accidents have resulted from low levels of controller SA and mode errors (Parasuraman & Byrne, 2002; NRC, 1997).

If controllers work OOTL for extended periods of time, their manual control skills may also degrade (NRC, 1998). This drawback of automation is known as skill decay (Endsley & Kaber, 1999). For example, in a highly automated system like CTAS or STARS (Standard Terminal Automation Replacement System), over time controllers may lose the skills required to perform basic ATC functions such as flight tracking or ground and aircraft close proximity detection (NRC, 1998). This becomes problematic when a situation requires a controller to revert to manual control, as previously mentioned. Recurring training, periods of manual control, and thoughtful automated system design are important for maintaining controller skills (NRC, 1998).

New, advanced forms of automation, including AA, are being considered for ATC to alleviate some of these disadvantages of conventional automation, and to preserve operator SA. The problems identified here can be attributed to operators acting as passive controllers, while removed from a primary system control loop. Adaptive automation design may provide a solution to these problems, as humans and computers can be effectively integrated in system

control loops in order to maintain controller SA and promote accurate decision making. In the following sections, we review contemporary research on AA in ATC and the SA implications of AA.

Adaptive Automation Research in the Context of ATC

In identifying forms of ATC automation, Hopkin (1998) suggested that automation cannot only be applied statically to different types of system functions at different levels, but it can also be dynamically applied to functions to switch control between the human and machine. Some recent research has explored the use of DFAs in the context of ATC simulations. One approach to AA in ATC is to introduce automation during periods of high workload (e.g., emergency procedures, high traffic density, aircraft non-conformance), when the controller may need assistance, and it can be shut-off during low workload periods (e.g., low traffic density) (Hilburn et al., 1997; Parasuraman, Wickens & Sheridan, 2000).

A number of studies have provided evidence that this type of approach to automation can be beneficial to radar monitoring or ATC tasks (Kaber & Endsley, 1997; Hilburn et al., 1997; Clamann, Wright & Kaber, 2002; Kaber, Prinzel, Wright & Clamann, 2002; Clamann & Kaber, 2003). Kaber and Endsley (1997) found that AA produced significantly superior performance than completely manual or fully automated control in a simple radar monitoring task. Hilburn et al. (1997) found similar results. They conducted a study to determine the effects of AA on decision making in ATC using the previously described CTAS DA. They developed three different automation schemes including constant manual control, constant automation and AA, which introduced automation during high air traffic periods. They found that the AA condition resulted in the smallest increase in workload compared to fully manual and automated control.

Kaber, Prinzel et al. (2002) and Clamann et al. (2002) also conducted research to describe the performance and workload effects of AA applied to a broad range of human-machine system information processing functions, including information acquisition, information analysis, decision making and action implementation, in the context of a low-fidelity ATC simulation. The forms of automation were similar to those presented by Parasuraman et al. (2000). Kaber, Prinzel et al. (2002) used a secondary gauge-monitoring task as a measure of primary task workload, which served as a basis for driving the adaptive automation. They found that humans are better able to adapt to AA when applied to lower-order sensory and psychomotor functions, such as information acquisition and action implementation, as compared to AA applied to cognitive (planning and decision making) tasks. Results indicated that operator performance was greatest when automation was applied to the action implementation aspect of the task. Results on secondary task workload measure indicated that performance was greater under both automated and manual control when AA was applied to the information acquisition aspect of the task. Automation of information acquisition appeared to relieve some task time pressure for subjects,

Another design parameter to address when incorporating AA into a system is the question of who will decide when automation will be invoked, or who has decision authority. Kaber and Clamann (2003) recently investigated the effects of different forms of invocation authority over DFAs of various information processing functions in a low-fidelity ATC simulation similar to that used by Kaber, Prinzel et al. (2002). Kaber and Clamann (2003) considered various types of

automation management authority, including human operator authority, computer authority, or authority distributed between both servers through mutual suggestions and approvals. Adaptive automation was applied to four stages of human-machine system information processing, including information acquisition, information analysis, decision making and action implementation, as part of the ATC simulation. This study also used a workload-based approach and facilitated DFAs through two levels of computer authority (suggestion and mandate). Results confirmed performance differences due to AA across the various aspects of information processing as well as between the two types of invocation authority (Clamann & Kaber, 2003). Subjects performed significantly better in the primary task during periods of automation as part of information acquisition AA as compared to decision making. They also performed significantly better when automation was suggested as compared to mandated.

The results of these research studies provide evidence that AA may improve ATC performance over completely manual control and static automation. They also indicate that the effectiveness of AA in the context of ATC may be dependent upon both the type of automation presented to an operator and the type of invocation authority designed into the system. As research and development make forms of AA more accessible and expose the advantages of using new methods, they may become actual approaches for overcoming some of the disadvantages of existing conventional automation in ATC.

Research on Measures of SA and AA

A review of the literature on existing approaches to SA measurement was conducted as a basis for structuring a new measure for application to adaptive systems. Many measures of SA have been developed over the past 10-15 years, including direct, objective measures such as the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1995b). SAGAT involves comparing an operator's perceptions of a task environment to some "ground truth" reality. This is accomplished by freezing a simulation exercise at random points in time and hiding task information sources (e.g., blanking visual displays) while subjects quickly answer questions about their current perceptions of the simulation. The questions are based on a cognitive task analysis of the simulated operations and are used to determine what the participant knows or comprehends about a scenario at the time of the operation. Subject responses to questions are then compared to actual data on the real situation collected by the computer system running the simulation in order to provide an objective measure of SA. The key drawbacks of this method include the need to temporarily halt a simulation exercise and the potential for increasing participant workload through queries and altering task performance (Endsley, 1995b).

Other direct, objective measures of SA include real-time probes, or queries posed directly to operators during task performance. Jones and Endsley (2000) studied the use of real-time probes as a measure of SA in complex system operations in order to determine the sensitivity and validity of such a measure in comparison to existing SA measurement techniques, including SAGAT. The probes that Jones and Endsley (2000) posed to operators were based on a SA requirements analysis for the system they studied. They verbally administered the probes (one-at-a time) during task performance (military radar monitoring under normal and simulated wartime conditions). Results demonstrated probes to be sensitive to workload manipulations in terms of overall response time and the accuracy of responses. With respect to the validity of the

real-time probes for assessing SA, Jones and Endsley (2000) found significant, but weak correlations of response accuracy with the accuracy of responses to SAGAT queries, which were also found to be a sensitive measure of SA. Although these results are promising for the use of a real-time probe measure in adaptive systems, it is possible that probes may be obtrusive to operator performance in that questions must be posed directly, during tasks. That is, probes may demand attentional resources subtracting from operator cognition on the test task. Beyond this, the validity of this approach to tapping the construct of SA has yet to be confirmed.

Previous AA research has demonstrated SAGAT to be sensitive to dynamic changes in system states (automation states) (Kaber & Endsley, 2004), as well as changes in adaptive interface content over time (Kaber, Wright & Hughes, 2002). Kaber and Endsley's (2004) recent research supports this contention. Their study evaluated the performance and SA effects of various forms of complex system automation adaptively allocated during operator performance in a simple version of the dual-task scenario used by Kaber, Prinzel et al. (2002). Manual and automated control allocations in the primary task occurred on the basis of pre-programmed schedules. Kaber and Endsley (2004) found that the level of task automation played a significant role in operator SA measured using SAGAT, as compared to the schedule of control allocations. Specifically, results demonstrated that intermediate levels of automation, including computer assistance in the planning and implementation aspects of the task (shared control), produced the highest operator comprehension of system states. However, SAGAT performance was the worst under the batch processing mode as well as high-level automation (supervisory control). Kaber and Endsley (2004) also observed on the basis of the SA results that operators were more able to deal with DFAs of levels of automation primarily applying computer assistance to decision-making aspects of the dynamic control task, as compared to levels applying automation to monitoring and implementation roles. This research also suggests that the impact of AA on SA may be dependent upon the human-machine system information processing functions to which AA is applied. Although lower forms of automation, such as action implementation aiding, may promote performance when adaptively applied to complex systems, they may also have a negative effect on SA. These findings should be of concern to designers of AA systems because it may mean that applying AA to psychomotor functions to increase performance in the near term may undermine SA over extended periods.

Related to this research, Kaber, Wright and Hughes (2002) recently studied the SA effects of AA of a complex remote-control robotic system (a telerobot) and the effectiveness of various forms of system feedback on DFAs for maintaining operator SA. In this experiment, manual and automated control (supervisory control) allocations during task performance were based on critical task events (e.g., transitioning from a robot search mode to a demolition mode). Kaber, Wright and Hughes (2002) observed AA-induced decrements in SAGAT performance under both feedback (visual and auditory cues) and no feedback conditions; however, sensory cues on DFAs did improve overall operator SA in comparison to no cues whatsoever. Consequently, if AA is applied to complex system functions, and designers are concerned about potential implications on operator SA and decision making, feedback on system state may be used to promote performance, particularly in cases in which lower forms of automation (e.g., information acquisition, action implementation) have been developed.

Considering this research, and the findings of Jones and Endsley (2000), the present study used the SAGAT measurement technique to describe the implication of AA on operator SA, performance and workload in ATC. In general, exploration of operator SA as a basis for triggering DFAs in adaptive systems would substantially extend current research on AA. It is possible that a SA trigger of DFAs may interact with the application of automation to various information processing functions in a different manner than a workload-based trigger. In this research, we also investigated the SA effects of AA applied to various information processing functions. The following sections present the specific aims and experiments as part of this project.

2. Objectives and Overview of Research

The objectives of this research included:

- (1) development of an enhanced version of the ATC simulation used by Kaber and Clamann (2003) for the human factors experiments as part of this research;
- (2) definition of a SA measure sensitive to automation state changes in adaptive systems;
- (3) empirical assessment of the SA measure for use in investigating the effectiveness of AA of various ATC information processing functions; and
- (4) description of the influence of AA in an ATC simulation on operator SA.

With respect to the experimental objectives, this study assessed the performance, SA and workload effects of AA of four different stages of human information processing in an ATC simulation, including information acquisition, information analysis, decision making and action implementation.

Enhancement of the Multitask© Simulation

Two previous versions of the Multitask© simulation were developed for studies of the performance and workload effects of AA applied to various stages of information processing in ATC (Kaber, Prinzel et al., 2002; Kaber & Clamann, 2003). As part of the current research, we significantly modified the Multitask© simulation in order to promote the level of realism of the task. The major enhancements to the simulation over earlier versions included the addition of multiple airports within the sector of airspace being managed by the operator, the addition of multiple holding fixes about each of the two airports, and the capability for operators to issue speed clearances, to command holding patterns, to re-route aircraft from one airport to another, and to advise of runway changes.

The new Multitask© interface is presented in Figure 1. It includes a radarscope, control and status boxes (see left side of figure) and a menu bar. The menu bar is used to select experimental trial settings, skill levels, and automation modes, all of which will be discussed in later sections. The radarscope represents approximately 15,400 square nautical miles (nm) of airspace within the user's control, and is separated into four quadrants by a horizontal and vertical line, each delineating the north, south, east, and west cardinal directions on the display. Concentric circles are presented on the display, each representing 10 nm increments from the display center. Near the center of the radarscope are the two airports, one 10 nm west of the center and another 10 nm

east of the center (with 20 nm between the airports). Each airport has two runways. Eight equally spaced holding fixes are also represented on the display by small circles located 30 nm from the center of the radarscope.

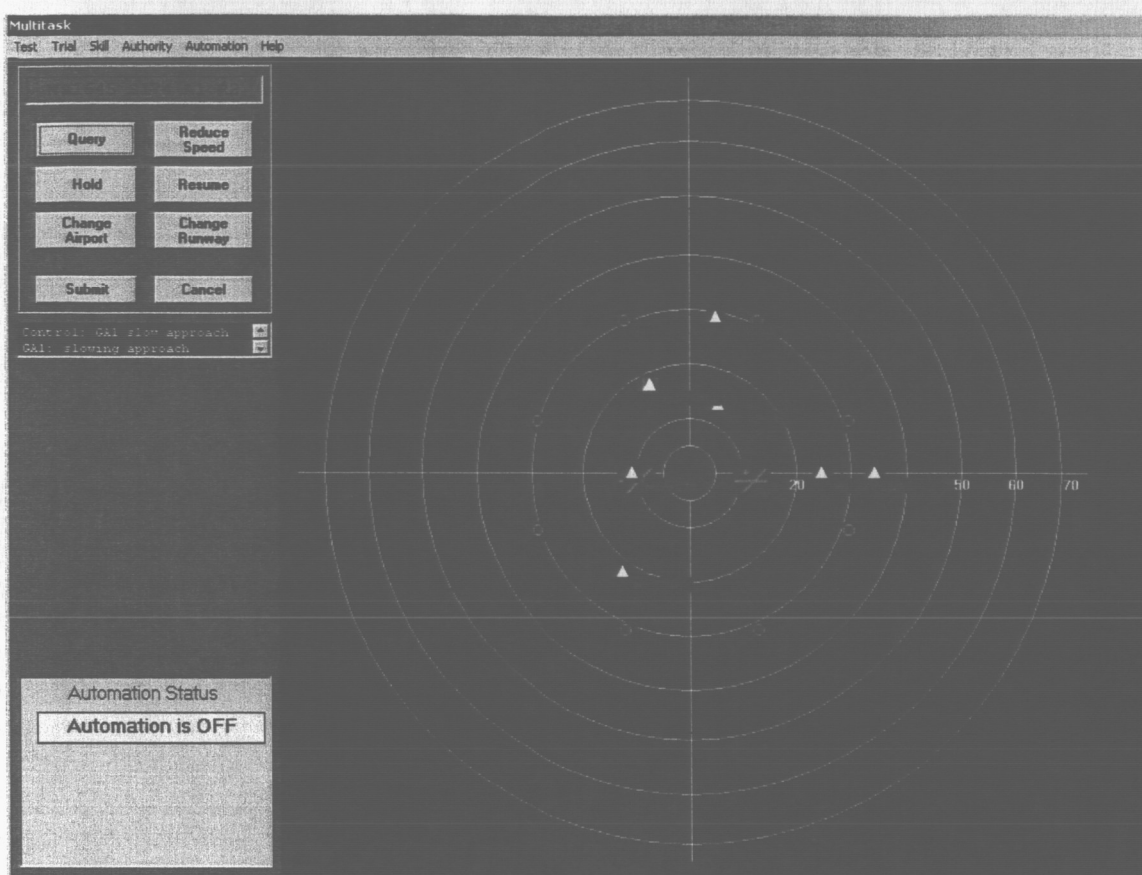


Figure 1. Multitask© display in manual control mode.

Simulated aircraft are represented on the display by triangle icons and data tags including the aircraft's call sign (see Figure 2). The aircraft data tag is similar to those found on ATC radar displays. During simulation run time, aircraft first appear toward the perimeter of the display on one of eight approach trajectories and move toward one of the two airports, destined for one of the two runways at an airport. (During the second experiment as part of this research, aircraft initially appeared at the edge of a predefined control sector (see the hexagon overlay on the radarscope near the 50 and 60 nm bands), which reduced the total area of control for subjects. It also served to accelerate the action of the scenario). The aircraft starting location (within the control sector), airport, and runway are all randomly assigned, and the aircraft appear one at a time. In the first experiment as part of this project, a buffer of 1-20s was used to delay the creation of each new aircraft, and in the second experiment a fixed buffer of 25s was programmed as part of the simulation. (It was observed during the first experiment that short buffer times created potential aircraft conflict situations at the periphery of the control sector, which were very difficult for subjects to address.)

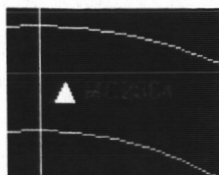


Figure 2. Aircraft icon and data tag.

When an aircraft first appears on the scope, triangle icon is white and flashing, indicating that the aircraft has not yet been contacted. Once an aircraft is contacted, the icon becomes a solid white triangle, and the aircraft's predetermined clearance (including call sign, aircraft speed, destination airport, and destination runway) is displayed in a data box on the left side of the Multitask© display (see upper-left corner of Figure 1). If an aircraft is placed in a holding pattern, the icon becomes yellow until the aircraft is advised to resume its approach to the destination airport, at which time it becomes white again.

The aircraft icons represent one of three possible aircraft types: commercial, private, or military. The aircraft call sign is presented as an alphanumeric designator, including a letter for aircraft type and a number, as shown in Table 1. The type of aircraft also dictates the possible range of speed for the vehicle. The exact speed is determined based on a standard amount of time required for each aircraft to reach its destination (without a controller clearance amendment). Table 2 presents the average time and speed values (in real-time) for each type of aircraft. The table also indicates the frequencies at which each aircraft type occurs in the simulation. The destination airports include A1 for the West airport or A2 for the East airport, and the runways include R1 or R2.

Table 1. Aircraft designators.

Aircraft Type	Letter designator	Numeric designator
Military	M	Sequential number
Commercial	Random 2-character airline designator	Random 4-digit number
Private	GA	Sequential number

Table 2. Aircraft type parameters.

Aircraft Type	Travel time	Average speed	Frequency
Military	21 minutes	200 kts.	10%
Commercial	25 minutes	170 kts.	70%
Private	42 minutes	100 kts.	20%

The control box, as part of the Multitask© interface, allows the user to communicate with all aircraft currently on the radarscope and to request changes to aircraft parameters. It includes the previously mentioned data box, eight command buttons, and a history box (see Figure 3). The data box is located at the top of the control box. The information in this box is displayed until another aircraft is contacted. The eight control buttons facilitate five clearance change commands, including reduce speed, hold, resume, change airport, and change runway, as well as two action commands, submit and cancel. The 'Query' command is used to initiate communication with an aircraft and obtain the aircraft's flight identifier and parameters. The

'Reduce Speed' command simply reduces the speed of an aircraft by a set amount, and the original speed cannot be reassigned. The 'Hold' command is used to request that an aircraft fly directly to the nearest holding fix and remain there for either 30 minutes or until advised to continue to the assigned airport and runway (the 'Resume' command). The 'Change Airport' and 'Change Runway' commands advise aircraft to change from their randomly assigned clearance to the alternate of the two possible destinations. The 'Submit' button must be used after selecting a desired clearance command, and 'Cancel' may be used to prevent an aircraft from further processing a clearance command. Controllers use all of the command keys to prevent possible conflicts between aircraft while maintaining landing efficiency (e.g., issuing the fewest number of amendments to original clearances). The history box, located below the command buttons (see Figure 3), displays the actual communications with an aircraft in text form, simulating actual ATC communications.

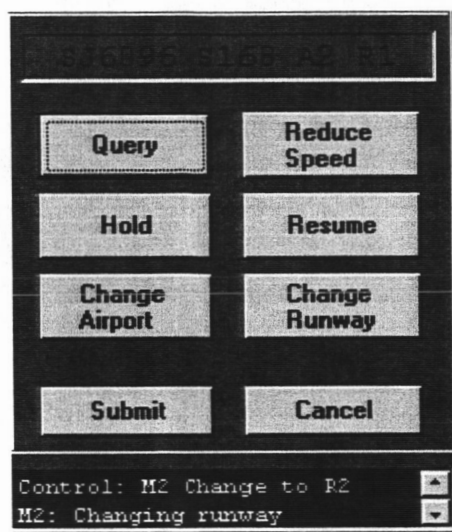


Figure 3. Multitask© control box.

Beneath the command box is an automation aid display box (see left-center of Figure 4). This box is used to present information to a controller that is pertinent to the current level of Multitask© automation. The aid is inactive under the manual control (no automation) of the simulation and one form of low level automated control. The information presented in this box under all other modes of Multitask© is described below.

An automation status monitor is located below the automation aid box and indicates whether automation or fully manual control is currently being utilized (see left-bottom of Figure 4). During AA trials when the computer intervenes in human operator control, the automation status control box displays, "Automation is ON," and during manual control portions of AA trials, the box displays, "Automation is OFF."

Multitask© is capable of operating under one of the following five modes of automation during a simulation trial:

1. Manual control – No assistance is provided. Operations are performed as described in the previous sections.
2. Information acquisition – A scan line rotates clockwise around the radar display, and as it passes over an aircraft icon, a TPA for that aircraft is presented for 2 seconds (see Figure 4 and TPA for Aircraft GA2). The TPA shows the aircraft destination and route in the form of a line connecting the vehicle and the airport or holding fix. The aircraft speed (in knots), destination airport, and destination runway are affixed to the center of the TPA for the aircraft (see lower-right quadrant of radarscope). (The automation aid box is inactive under this mode.) This form of automation essentially allocates the sensory processing stage of ATC information processing to machine control.

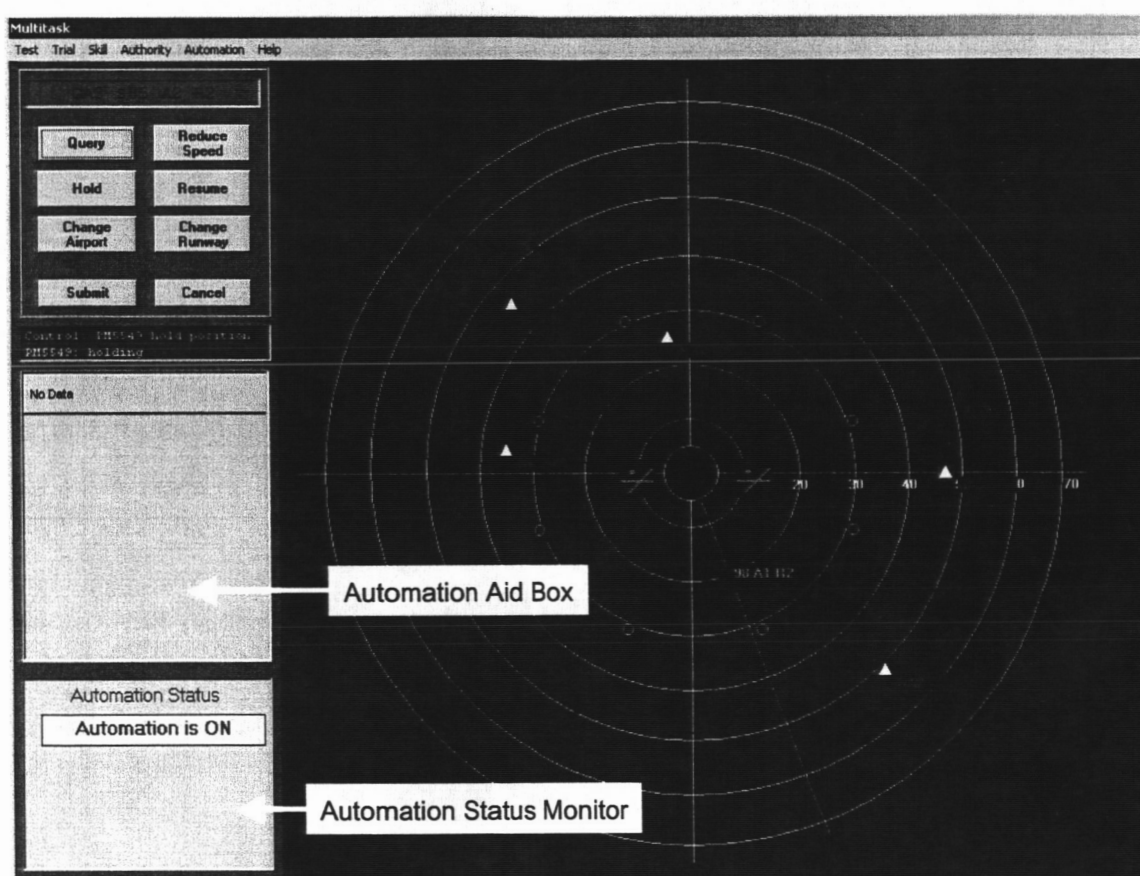


Figure 4. Information acquisition automation display screen.

3. Information analysis – Information pertinent to each of the contacted aircraft on the radarscope is displayed in a table in the automation aid box, including the aircraft's call sign, destination airport, destination runway, speed, and distance (nm) from the destination airport (see Figure 5). A final column, 'Conf,' denotes the call sign of any aircraft that are currently in conflict with each other. This form of automation assists operators with the integration of perceived information and long-term memories as part of information processing.

ID	A	R	Spd	Dist	Conf
SJ3455	1	1	SPD1	54.7	
DK3955	2	1	SPD1	54.5	GA1
MS5580	1	2	SPD1	42.7	MC3365
RM3368	1	2	SPD1	16.7	
MC3365	2	2	SPD1	29.1	MS5580
GA1	2	1	SPD1	62.6	DK3955

Figure 5. Information analysis automation aid box

4. Decision making – In addition to conflict alerting, recommendations for conflict resolution are provided in a table in the automation aid box. Information on conflicting aircraft, the recommended clearance change (speed, hold, resumption, airport, or no option), and which aircraft to advise of a necessary change are all displayed together (see Figure 6). Up to three automated clearance recommendations are displayed in the automation aid box at a time and they are listed in order of priority. This form of automation assists operators with information processing requirements associated with the decision and response selection aspects of the task.

FLIP - SPEED - DIST			
SJ6933	DK2668	Reduce Speed	SJ6933
MS1383	GA4	No Option	MS1383

Figure 6. Decision making automation aid box.

5. Action implementation – This form of automation simulates the “hand-off” of aircraft control from approach control to local-tower control, and the tower automatically maintains full control responsibility for aircraft within 20 nm of the center of the radarscope. This type of automation prevents any conflicts after “hand-off” to tower control. The action implementation display includes a table that summarizes the classification and number of aircraft on the display (see Figure 7). Action implementation

automation assists the operator with the requirement of response execution as part of the ATC simulation.

Class	Total	
Military	1	
Private	2	
Commercial	2	

Figure 7. Action implementation automation aid box.

Under all modes of automation, the objectives of the controller in the task are to contact aircraft appearing on the radar display, make any necessary changes to pre-existing aircraft clearances based on their potential to cause a conflict, and safely land aircraft at one of the two airports. Multitask© performance is measured in terms of the number of aircraft cleared, the potential collisions (conflicts), actual collisions, and the number of clearance amendments administered. This data is recorded during simulation trials and displayed at the end of the task (see Figure 8). Aircraft arriving safely at an airport are considered cleared aircraft. Aircraft traveling within 3 nm of other aircraft, as they travel to their assigned airport, or two aircraft that are within 20 nm of the center of the radarscope, destined for the same runway at the same airport, are considered potential collisions. (During the second experiment as part of this study, when subjects used decision making or information analysis automation, auditory alerts were provided to warn operators of these potential collisions. It was observed during the first experiment that some potential aircraft conflicts were not salient to subjects based on their understanding of the conflict criteria. The use of auditory cues on conflicts was in-line with the theoretical description of the decision making and information analysis forms of automation provided by Parasuraman et al., 2000.) Aircraft that simultaneously arrive at the same airport destined for the same runway, or aircraft that come in contact with each other, constitute actual collisions in which case auditory feedback is provided and both aircraft are removed from the screen. In addition, the number of clearance changes is recorded as a measure of performance. In general, operators attempt to minimize the number of clearance changes issued. (More details on the simulation performance measures are presented in the description of the independent and dependent variables as part of the human factors experiments.)

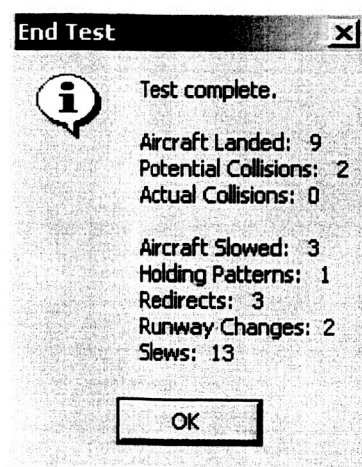


Figure 8. Multitask© performance measures

The experimenter controls the form of automation applied to the Multitask© simulation, the speed of the simulation, the duration of trials, and the skill level. During experimental trials, the various modes of automated assistance can be switched “on” or “off”, or adaptively applied to the task, based on operator workload states. However, only one mode can be used per trial. The Multitask© skill level is defined by the number of aircraft (from 1 to 9) that are present on the radarscope at any given time. In general, the Multitask© features and options described in this section represent substantial enhancements over the previous versions of the software in terms of task complexity and realism relative to actual TRACON.

Defining a Measure of SA Sensitive to Dynamic Function Allocations as part of AA

We defined a new SA measure based on a review of previous research, involving goal-directed task analyses (GDTA) of ATC operations, and by applying the GDTA methodology to the enhanced Multitask© simulation. Endsley and Rodgers (1994) and Endsley and Jones (1995) conducted research to determine the goals and critical decisions in ATC tasks, as well as controller SA requirements to achieve ATC goals. Both studies used knowledge elicitation techniques with experts and applied the GDTA methodology. Endsley and Rodgers (1994) interviewed retired air traffic controllers and analyzed simulated ATC scenarios to develop a very extensive list of SA requirements for en route controllers. (En route ATC refers to the control centers across the country responsible for navigating, separating, and “handing-off” aircraft between airports, or terminal air space (Nolan, 1999).) The en route ATC analysis consisted of overarching ATC goals, subgoals required for accomplishing the overarching goals, and critical decisions that must be addressed by the controller in order to accomplish subgoals. (These decisions can ultimately be used as a basis for developing SA queries in order to assess controller perception, comprehension and projection in tasks.) Finally, the high-level and low-level SA requirements for addressing the decision (answering questions) were identified by Endsley and Rodgers (1994).

Endsley and Jones (1995) performed a similar analysis to evaluate the SA requirements for TRACON centers. A TRACON area refers to the controlled airspace within the immediate vicinity of a congested airport (Nolan, 1999). This work has greater relevance to the present

study, as the Multitask© simulation is a low-fidelity representation of a TRACON scenario. A portion of one GDTA conducted by Endsley and Jones (1995) for avoiding aircraft conflicts during TRACON operations (i.e., separate aircraft) is presented here with labels differentiating the various elements of the GDTA:

- [goal] 1.1 Separate aircraft
- [subgoal] 1.1.1 Assess aircraft separation
- [question] - *vertical separation meets or exceeds limits?*
- [high-level SA information] -- vertical distance between aircraft along route (projected)
- [high-level SA information] -- vertical distance between aircraft (current)
- [low-level SA information] -- aircraft altitude (current)
- [low-level SA information] -- altitude accuracy
- [low-level SA information] -- altitude (assignment)
- [low-level SA information] -- altitude rate of change (climbing/descending)

(Endsley & Jones, 1995, p. 19)

In general, this type of analysis has been established as an important tool for understanding ATC SA requirements, and is valuable for designing and developing future ATC systems by considering controller information needs.

The GDTA methodology was applied to the Multitask© simulation, following the approach taken by Endsley and Jones (1995). The major goals of the task and subgoals were broken-down to identify decisions that must be made to accomplish the goals. As an example, a portion of a GDTA developed for the Multitask© subgoal, "acquire aircraft information," is presented here:

- 1. Acquire aircraft information
 - 1.1 Locate aircraft
 - in what display sector is the aircraft located?*
 - aircraft position
 - how many other aircraft are located in that sector?*
 - aircraft position
 - location of other aircraft

Identification of the critical decision for this subgoal, including, "in what display sector is the aircraft located?" led to identification of the perception, comprehension, and projection requirements for accomplishing "Locate aircraft", etc. in the context of the Multitask© simulation. These decisions and SA requirements were used to develop SA queries as part of the new measurement technique. The complete GDTA for the Multitask© simulation is presented in Appendix A.

As previously mentioned, we elected to use a SAGAT-based approach to SA measurement in this study because of its demonstrated validity and reliability, and the potential drawbacks of real-time probes in terms of task disruption and operator reallocation of attentional resources. Situation awareness queries are a key aspect of the SAGAT methodology. Examples of Level 1 (perception), Level 2 (comprehension), and Level 3 (projection) SA queries developed for this

research are presented in Table 3. All queries are categorized according to the levels of SA defined by Endsley (1995a). The complete list of SAGAT queries developed for this research is presented in Appendix B.

Table 3. Example SAGAT questions.

Level 1 SA	What is the aircraft's call sign?
Level 2 SA	What is the distance of the aircraft from its destination (in nm)? (Criterion: Subject's answer must be within 5 nm of actual situation for it to be graded as correct.)
Level 3 SA	When will the aircraft arrive at the destination airport (in min. from now)? (Criterion: Subject's answer must be within 2 min. of actual time for it to be graded as correct.)

With respect to our approach to implementing the SAGAT measure, we followed Endsley's (1995b) original methodology, including administering SA queries to operators during simulation trials and task freezes. Level 1 SA queries often constitute a memory test for subjects and typically they are presented with a basic "map" of the work environment and asked to locate various elements in the task scenario (Jones & Kaber, in press). For example, in Kaber and Endsley's (1997) research a graphic of their radarscope display along with numbers representing individual aircraft were presented to subjects and they were asked to recall specific aircraft attributes based on the location information provided. That is, the graphic, or key, formed the basis for the questions administered during simulation freezes. As subjects answered questions about specific aspects of the task, a computer or experimenter records the "ground truth" of the simulation scenario. Subsequent to responding to the questions, the simulation was resumed for the subject.

Endsley (1995b) identified several guidelines for implementing SAGAT including:

1. The timing of the freezes should be random, in order to prevent subjects from anticipating a query and to ensure queries do not occur during the same phases of a simulation or that they only occur during high activity.
2. The queries should not occur within the first 3-5 minutes of a trial to allow a subject time to develop SA.
3. Two queries should not occur within 1 min. of each other to allow subjects time to reacquire SA after a freeze.
4. Multiple freezes during a trial are appropriate. Endsley (1995b) found that 3 stops during a 15-min trial did not adversely affect subject performance in a simulation.
5. An entire experiment should collect 30-60 SA samplings per experimental condition.

Responses to SAGAT queries are scored as either correct or incorrect. Jones and Kaber (in press) said that the responses to each query are typically analyzed separately; that is, all of the responses to one query are analyzed across conditions. Many researchers have effectively used this approach to SAGAT data analysis including the evaluation of pilot SA with aircraft cockpit systems (Endsley, 1995c), controller SA in ATC (Endsley & Kiris, 1995), and radar monitors in the theater of defense tasks (Bolstad & Endsley, 2000). Other recent research on SA in ATC (Hauss & Eyferth, 2003) has also demonstrated that analyzing SA questions individually can be an effective method for assessing controller SA. The results for each query are then generalized

to the appropriate SA Level, and the scores for each experimental condition serve as a direct indicator of controller SA during those conditions.

Jones and Kaber (in press) also describe a method of analysis in which SAGAT queries are categorized according to the levels of SA defined by Endsley (1995a) and composite scores for Level 1, 2, and 3 SA are computed based on the accuracy of subject responses to the sets of questions. This was the primary method of SA data analysis used in the experiments as part of this research in order to assess the impact of various forms of AA on air traffic controller performance and situation awareness.

The key advantage of the SAGAT methodology is that it provides diagnostic information about specific elements of operator SA, which can be collected during a trial without increasing operator workload (Endsley, 1996; Jones & Kaber, in press). Applying this method to the study of human-automation interaction can provide a clearer understanding of how ATC systems design and AA, in particular, may influence user SA. The specific implementations of the SAGAT methodology in the experiments as part of this research are detailed below.

Quantifying the impact of AA on SA in an ATC simulation

Two experiments were conducted as part of this research. The first experiment was designed to validate the measure of SA for sensitive and reliable assessment of operator perception, comprehension and projection in the ATC simulation. We wanted to see if the measure would reveal significant differences in levels of operator SA during manual versus automated control of the Multitask© simulation. It was generally expected that operator SA would be superior under manual control, when operators were playing an active role in aircraft clearances.

The second experiment was to assess the SA implications of AA applied to the various stages of Multitask© information processing, and to evaluate an SA-based approach to triggering DFAs as part of AA. We wanted to determine if changes in SA occur in association with DFAs, based on operator workload, and the magnitude of any changes. On the basis of the results of prior research (Kaber & Endsley, 2004), it was generally expected that the SA measure would be more sensitive to automation manipulations impacting higher-order information processing functions, including information analysis and decision making. (Unfortunately, the latter objective for the second experiment was not achievable due to findings on the SAGAT measurement approach obtained through the first experiment. This is explained in detail in the Results and Discussion sections.)

3. Experiment #1

Participants

The first experiment as part of this project involved a sample of eight subjects on which repeated measures of performance, SA and workload were collected during AA of the enhanced Multitask© simulation using a workload-based approach to DFAs. Subjects ranged in age from 23 to 27 years. All subjects had 20/20, or corrected to normal, vision and personal computer experience. None of the subjects had flying or ATC experience. Subjects were compensated at a

rate of \$10 per hour for their participation. At the beginning of the experiment, they were also informed of an additional incentive, a \$30 gift certificate, which was awarded to the top performer.

Tasks

The experiment used a dual-task scenario similar to that employed by Kaber, Prinzel et al. (2002) and Kaber and Clamann (2003) involving subject performance of the enhanced Multitask© simulation (as described above) and a secondary, gauge-monitoring task to objectively assess operator workload. The gauge task included a fixed scale, moving pointer display with a central “acceptable” region bordered on either side by two “unacceptable” regions (see Figure 9). The acceptable region was colored “green”, the unacceptable regions were colored “red”, and two small transitional areas were colored “cyan”. The transitional ranges were also considered acceptable gauge values.

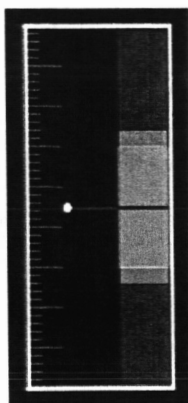


Figure 9. Gauge-monitoring display.

The user's goal was to detect and correct pointer deviations into either unacceptable region by using keys on a keyboard ('Shift' to move the pointer up or 'Ctrl' to move the pointer down), depending on which direction the pointer drifted. After returning the pointer to the acceptable range, it would continue to drift randomly until the end of the test trial.

Errors in the gauge-monitoring task were defined as pointer deviations into either of the unacceptable regions (signals) without a control response from the operator to correct the deviation (a miss). Errors were also recorded when operators commanded an unnecessary correction (a false alarm; e.g., a 'Ctrl' key press when the pointer was in the green or blue ranges).

The gauge-monitoring task was programmed to prevent the randomly moving pointer from dwelling in an unacceptable region for more than 3s at a time. Clamann et al. (2002) determined 3 seconds to be an adequate amount of time for subjects to register a deviation while operating the primary ATC simulation task. The design of the gauge task allowed for approximately six pointer deviations, or signals, to occur per minute. Gauge task performance was recorded as a hit-to-signal ratio, or the number of times the subject detected the pointer in an unacceptable region, divided by the number of signals.

Experiment Equipment

The equipment used to present the tasks to subjects included a high-performance Dell (530) graphics workstation with 512 MB of RAM and dual 2 GHz Pentium Xeon processors. Multitask© and the gauge-monitoring task were presented on two 17-inch, digital flat panel displays (see Figure 10). Two keyboards and two, 2-D computer mice were integrated with the computer system. Both the subject and the experimenter used a keyboard and mouse during test trials. The subject's interface controls were used to control the Multitask© and gauge-monitoring tasks, while the experimenter's keyboard and mouse were used to facilitate simulation trial freezes. (Experimenter controls not shown in figure.)

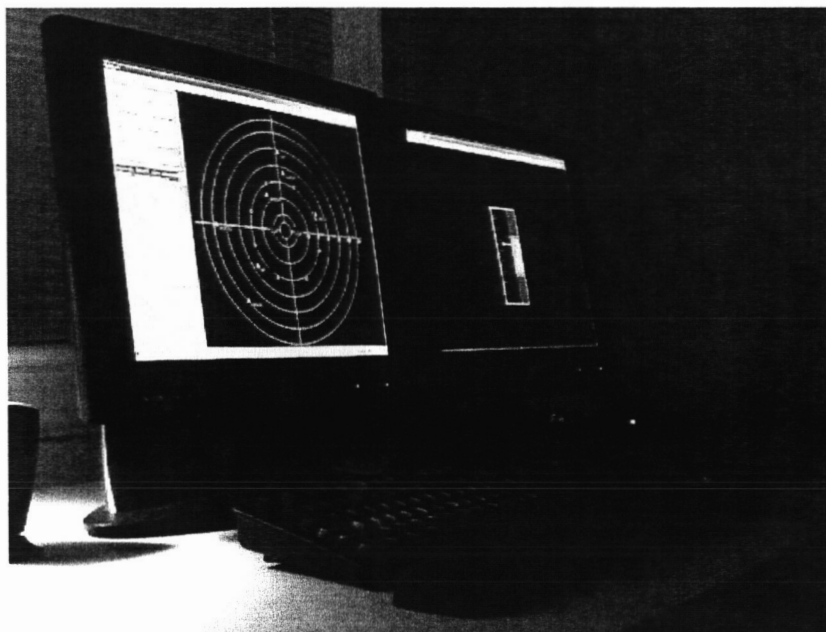


Figure 10. Multitask© and gauge monitoring task workstation.

A separate worktable and computer were set up adjacent to the Multitask© simulation workstation and were used to present SA questionnaires to subjects. The experimenter sat out of view of subjects and took particular care to not distract subjects while the simulation was running. Simulation freezes were periodically conducted to allow subjects to move to the second computer workstation and respond to the SA queries while the experimenter obtained correct answers to the queries from the Multitask© console. (The specific experimental procedures are discussed in detail, below.)

Approach to AA

The gauge-monitoring task provided an index of operator workload in the Multitask© simulation. A low score in the monitoring task implied a high level of workload in the ATC simulation and vice versa. Since the perceptual and cognitive demands of Multitask© functions overlap those of the gauge-monitoring task, previous research (Kaber & Riley, 1999) has found

the gauge-monitoring task to be a sensitive indicator of workload changes in the Multitask© simulation. The gauge task also served as a basis for triggering DFAs in the Multitask© simulation, specifically shifts from manual control to one of the forms of automated control and vice versa. When secondary-task (gauge) performance was poor (suggesting an increase in operator workload), the Multitask© simulation shifted from manual control to automated control. If operator secondary-task performance was good (suggesting a reduction in operator workload), the Multitask© simulation returned to completely manual control.

During the training sessions as part of the experiment, the average gauge-monitoring performance level and the standard deviation (SD) for the hit-to-signal ratio on pointer deviations was recorded. Pairs of subjects performed the gauge task while simultaneously using a randomly assigned mode of automation to control the ATC simulation. Their performance was used to establish Multitask© “overload” and “underload” conditions for all subjects exposed to the same mode of automation during test trials. When a subject performed an experimental trial, if gauge-task performance dropped below average practice performance minus 1 SD for one minute of the trial (implying an increase in subject workload), the computer mandated a switch from manual control of the Multitask© to the mode of automation for the trial. Once the automation was initiated, if subject performance in the gauge task increased above the mean plus 1 SD, or a hit-to-signal ratio of 1.0 was achieved (perfect performance in the gauge task), for one minute of a trial, then manual control was restored. These criteria were developed by Kaber and Riley (1999) for a closed-loop AA system.

Experimental Design

The experiment followed a completely within-subjects design with blocking on the subject (i.e., participant exposure to experimental conditions was randomized). All eight subjects completed two, 30-min. trials under each of the five modes of Multitask© control. The design was replicated to allow for estimation of experimental error attributable to the subject. Multiple trials involving completely manual control or information acquisition, information analysis, decision making and action implementation automation, were used to determine whether the SAGAT was a sensitive indicator of SA in the simulated ATC task. The design allowed each subject to be used as a basis for comparison of the mode of automation effects on the SA response measure and performance, etc. and controlled for the variability among subjects (Montgomery, 2001, p. 127). It is important to note that each AA trial consisted of both manual and automated control periods. The number of manual and automated minutes during a trial depended on subject workload fluctuations that occurred during the ATC simulation (as indicated by gauge-monitoring performance).

Variables

The independent variable (IV) manipulated in this experiment was the mode of AA applied to the Multitask© simulation, including completely manual control, information acquisition automation, information analysis automation, decision making automation, and action implementation automation. Under the AA conditions, each mode of automation switched “on” and “off” depending upon operator workload states.

The dependent variables included Multitask© performance, which was recorded in terms of cleared aircraft, conflicts, and collisions. Another dependent variable was gauge-monitoring task performance, recorded as a hit-to-signal ratio. Both of these measures were captured during each minute of a test trial.

Percent correct responses to SA queries were recorded during the experiment using the SAGAT-based approach introduced above. Each trial incorporated 3 SAGAT freezes. The freezes were designed to occur within one of three 8-min windows of time (7 to 14-min, 15 to 22-min, and 23 to 30-min) to ensure a sampling of SA throughout testing. (A random number generator was used to generate freeze times within each of the windows.) The first six minutes of a trial were not considered for freezes in order to allow subjects time to acquire SA (Endsley, 1995b), and to allow the aircraft time to move from the periphery of the radarscope towards the airports. Furthermore, no two freezes were scheduled within 2-min. of each other in order to allow subjects time to reacquire SA after a freeze (Endsley, 1995b).

During each freeze, the Multitask© display was hidden, and subjects were posed with SA queries randomly selected from a pool of 18 total questions targeting the three levels of SA defined by Endsley (1995a; see Appendix B). Each freeze included 6 questions. At the beginning of a freeze, subjects were asked to identify the current locations of aircraft by marking-up a graphic of the radarscope with a pencil. Subsequently, they were asked to respond to each of the 6 SA queries for each aircraft by completing tables. The questions were administered electronically using a computer database application.

Since Multitask© is a truly adaptive system in which a real-time workload measure is used to trigger DFAs between manual and automated control, it is not possible to predict at which moments during a simulation trial a subject will be using manual or automated control. Consequently, it was not possible to pre-determine a distribution of SAGAT stops during trials that would produce an even number of SA queries under manual and automated control conditions. However, since the SAGAT freeze times were randomly determined, there was an equal likelihood that the SA queries were posed during Multitask© manual control periods or automated control periods. Since a large number of SAGAT freezes were conducted during the experiment, the average percent correct responses to SA queries under manual control or automated control was expected to be representative of actual operator SA.

Procedures

The procedures for the experiment are summarized in Table 4. Each subject began the first day of the experiment with an introduction and equipment familiarization period. An informed consent form and participation payment form were reviewed and signed by subjects during this period. Subjects were also shown the computer system monitors, keyboard, and mouse to be used during the simulation.

Table 4. Experiment timetable.

Steps on Day 1	Time
Introduction and equipment familiarization	15 min.
Multitask© training under manual control	15
Gauge-monitoring training	5
Information acquisition automation training	15
Information analysis automation training	15
Decision making automation training	15
Action implementation automation training	15
Break	10
Adaptive automation training	15
Dual-task (Multitask© and gauge-monitoring) practice*	25
SAGAT familiarization	10
Dual-task and SAGAT practice*	50
Subtotal:	205
Steps on Day 2	
Simulation Review	15
Experimental trial 1	50
Experimental trial 2	50
Break	10
Experimental trial 3	50
Experimental trial 4	50
Break	10
Experimental trial 5	50
Subtotal:	285
Steps on Day 3	
Simulation Review	15
Experimental trial 6	50
Break	10
Experimental trial 7	50
Experimental trial 8	50
Break	10
Experimental trial 9	50
Experimental trial 10	50
Debriefing	10
Subtotal:	295
Total duration for experiment:	785

* Subjects were randomly assigned to modes of automation with equal numbers of subjects (n=2) experiencing each mode.

Subsequently, a training and practice session was conducted. First, Multitask© and gauge-monitoring training was conducted. Afterwards, participants underwent training sessions on each of the 4 modes of automation and the AA function of Multitask©. This training included descriptions of the modes, screen familiarization, and a short practice trial for each mode of automation. The simulation training was followed by a dual-task practice trial in which both the Multitask© and gauge-monitoring task were performed under a randomly assigned mode of Multitask© automation. Subsequently, subjects were familiarized with the SAGAT, and sample questions were presented. Finally, a dual-task practice session, involving performance under the same randomly assigned mode of automation, was conducted with 3 SAGAT simulation freezes.

The SAGAT questions presented during the practice session were randomly selected from the questions to be posed during the actual experimental trials. Each subject performed the dual-task plus SAGAT practice trial under one mode of automation and the average gauge-monitoring task performance and SD were recorded.

During the second and third days of the experiment, subjects reviewed the simulation procedures and completed 10 experiment trials involving the use of the various modes of Multitask© automation. The order in which the modes were presented to subjects was randomized. Each trial lasted approximately 50-min., including 30-min. of simulation time and approximately 20-min. to answer SA questionnaires during all three freezes. All training, practice, and experiment sessions were conducted using the Multitask© skill level of seven (i.e., 7 targets were presented on the display at any given time). (The setting of seven targets was used based on the working memory capacity of 7, plus or minus 2, chunks, as defined by Miller (1956).)

All training and practice trials involved the ATC simulation running at 10-times above real time, with the exception of the dual-task and SAGAT practice sessions during which the Multitask© ran 2-times faster than real time. All test trials were also conducted at 2-times above real time. These simulation speeds were determined appropriate based on previous research conducted by Clamann and Kaber (2003), which successfully used similar accelerated speeds for training and testing. Furthermore, pilot testing suggested that 7 targets traveling at 2-times above real time was an appropriate level of workload for experiment trials.

The entire experiment lasted three weeks. Each subject was tested for approximately 13 hours over three separate days. The day-of-the-week and time of each subject's participation remained consistent during the experiment with the exception of two scheduling conflicts. At the conclusion of the third day of testing, each subject verified and signed the payment form, and they were debriefed.

Data Analysis

Performance, workload and SA response measures were calculated for both automated and manual control periods as part of AA trials. The experiment yielded 10 manual and 8 automated performance scores for each subject. Therefore, eighty manual performance observations were gathered across subjects through the completely manual control trials and AA conditions (8 subjects x 2 trials x 5 control modes). Only 64 automated performance scores resulted from the AA conditions (8 subjects x 2 trials x 4 automation modes).

Analyses of variance (ANOVAs) were conducted on all response measures (primary task performance, gauge-monitoring task performance, and SA) with the mode of AA and subject included as predictor variables in the ANOVA model. Type III sums of squares were used to determine the significance of the IV on responses (SAS Institute Inc., 1990, p. 120-21). The mean square error (MSE) for the specific statistical models was used in all F-tests on the IV. For a repeated measures design with a single treatment, and the subject term included in the statistical model, Montgomery (2001, p. 624) states that this is the appropriate form of the F-test and that such a model partitions the variance attributable to the subject. (A conservative

approach to assessing the significance of a treatment in human factors experiments is to use the mean square for the subject variable as a denominator in the F-test.)

Tukey's Honestly Significant Difference (HSD) tests were used to breakout the means associated with the various AA conditions and manual control condition. If there were significant differences in SA in the ATC simulation as a result of the modes of automation, Tukey's test was applied to the manual and automated control data to identify which mode of control produced superior results. An alpha level of 0.05 was used to identify any significant effect of the mode of automation. (The same criterion was also used in the second experiment as part of this project.)

Multitask Data

Two data points were removed from the Multitask© aircraft conflict data as the result of a subject failing to follow instructions during two trials (Subject 7, Trials 3 and 4). This reduced the number of observations on conflicts for the experiment from 80 to 78 and decreased the number of conflict observations on the manual and automated control periods during the AA trials from 64 to 63 (one of the trials was a manual control condition trial). Two data points were also removed from the collision data set as the result of an aircraft appearing on the radar display at exactly the same location as another aircraft (Subject 2, Trial 1; Subject 5, Trial 5). These two occurrences were considered to be simulation software errors.

Workload Data

Subject workload was objectively measured in terms of the hit-to-signal ratio in the gauge-monitoring task. Average ratios were calculated for each test trial (including completely manual control and automated control periods) and separate mean hit-to-signal ratios were calculated on performance during manual and automated control periods as part of the AA conditions. One observation was missing from the gauge-monitoring task data as the result of one (very short) automated control period not including any gauge pointer deviations (Subject 1, Trial 8). This decreased the number of observations for automation control periods to 63.

Diagnostics and Statistical Model

Subsequent to organizing the primary task and workload performance data into three data subsets (total performance, performance during automated control, and performance during manual control), data analyses were conducted using SAS. ANOVA model residual values were output and plotted against the model predicted values, mode of automation, and trial number to determine if the data met the assumptions of the ANOVA including random process, linearity, and constant variance. The normality assumption was also assessed using Shapiro-Wilk's test and normal probability plots. These diagnostics indicated if any transformations on the response or IV might be required to ensure the appropriateness of the parametric test, as prescribed by Neter, Wasserman, and Kutner (1990). Any transformations employed in the various analyses are described in the Results section. Based on the residual and trial number plots, no learning was found in either the Multitask© or gauge-monitoring performance data sets.

A statistical model in mode of automation and subject was used to assess the impact of the various AA conditions on response:

$$Y_{ij} = \mu + A_i + S_j + \epsilon$$

for

$$i = 1, 2, 3, 4, 5$$

$$j = 1, 2, 3, 4, 5, 6, 7, 8$$

where

Y_{ij} = Response measure

μ = Mean response value

A_i = Mode of automation

S_j = Subject

ϵ = Experimental error

Situation Awareness Data

On the basis of subject responses to the SA queries, average Level 1, 2, and 3 SA, and a total SA score, were determined for each freeze. The type of simulation control, manual or automation as part of the AA conditions, was also recorded for each freeze, and an analysis was conducted to identify any significant effects on SA. As previously mentioned, six of the 18 SAGAT questions were randomly posed to subjects at each SAGAT freeze. Therefore, 240 observations were expected for the entire experiment (8 subjects x 10 trials x 3 stops). Unfortunately, some subjects failed to comprehend certain queries or they skipped queries because they did not know the answers. In addition to this, some SAGAT freezes occurred in the last minutes of the simulation and posed questions for which the correct answers were indeterminable (see Level 3 SA Questions B and E in Appendix B). Finally, since SAGAT queries were selected at random for each freeze, some freezes did not include questions representing one of the three SA levels.

Subsequent to an analysis of the SAGAT data to determine any general control mode (manual or automated) effects, the percent correct responses to Level 1, 2, and 3 SA queries, and total SA for each trial were analyzed for potential effects of the specific AA conditions. The percent correct responses to each individual SA question were also analyzed for sensitivity to the automation manipulation.

Each SA question at each stop solicited 7 responses from subjects (a response was required for each of the 7 aircraft on the display); however, less than the 7 possible responses may have occurred during certain trials for the following reasons:

1. some subjects were not able to correctly identify all 7 aircraft on the graphic of the radarscope (see Appendix B) during simulation freezes;
2. some subjects did not respond to a particular question for all aircraft;

3. the correct answers for certain questions were not obtainable for some aircraft on the display (e.g., minutes until destination could not be projected when a trial ended with an aircraft in a holding pattern); and
4. some aircraft had not yet been contacted when freezes occurred; therefore, some information was not yet available to the subject (e.g., destination airport).

These instances necessitated that the SA responses be scored as a ratio of correct answers to possible responses (i.e., the number of correctly identified aircraft was used as the denominator for SA question scores).

Since the responses to SAGAT questions represent a binomial variable (correct or incorrect), the discrete nature of the responses violates the assumptions of the ANOVA. With this in mind, Endsley (1995b) found the arcsine function ($Y' = \arcsine(Y)$) to be an effective transformation to account for this problem in her use of SAGAT. Consequently, the arcsine function was applied to the percentages for each question, and the effectiveness of the transformation was verified using the same diagnostic plots and tests considered in evaluating the performance data.

Specific Hypotheses

Based on the results of prior research, we formulated a number of hypotheses on Multitask© performance, and operator workload and SA for testing through this experiment. It was expected that Multitask© performance would be worse during completely manual control as compared to any mode of automation (Hypothesis (H) 1). Clamann et al. (2002) observed that any AA of ATC information processing functions produced better performance than no automation whatsoever. Related to this, we expected superior Multitask© performance during trials in which AA was applied to lower-order sensory/response functions, such as information acquisition and action implementation (H2). Leroux (1993), Laois and Giannacourou (1995), and Kaber, Prinzel et al. (2002) all found that humans performed better in ATC simulations using automation providing assistance with sensory/response functions.

On the basis of Hilburn et al. (1997) research, we expected trials involving manual control to produce greater workload than trials involving automated control (H3). Based on Kaber, Prinzel et al. (2002) results, and observations made by Parasuraman et al. (2000), it was expected that higher levels of automation, including information analysis and decision making, presenting complex displays for operator interpretation might demand high levels of visual attention and actually increase workload (as evidenced by decreases in gauge-monitoring task performance) (H4). Related to this, we expected modes of automation not presenting large amounts of information for operators to process (e.g., action implementation), or presenting information directly on the radar display (e.g., information acquisition), to decrease workload (H5).

With respect to operator SA, as Endsley (1995a) and Wickens (2000) observed, high levels of automation may result in OOTL performance, and can produce lower levels of operator SA. Therefore, subjects were expected to do better at responding to SAGAT queries posed during the manual trials of the experiment and manual control minutes of AA trials than during periods of automation (H6). In general, subjects were also expected to perform better on SAGAT queries during modes of automation that maintained user involvement in the system, like information

acquisition (Kaber et al., 1999) (H7). With respect to action implementation automation, this mode automatically “handed-off” aircraft to tower control, eliminating the risk of a collision within 20 nm of the radarscope center. The mode was expected to improve Multitask© performance, but it also removed the controller from the control loop and was expected to negatively affect SA (H8). Level 2 SA query performance was expected to be worse under the information analysis LOA because that mode of automation removed operators from the process of integrating aircraft information in order to identify conflicts (H9). In addition, decision making automation trials were expected to decrease the accuracy of responses to Level 3 SA queries because operator involvement in the clearance decision process was limited, potentially inhibiting their ability to predict the future status of aircraft (H10).

Results and Discussion

Primary Task Performance

An ANOVA on Multitask© performance revealed no significant differences between the completely manual control condition and the modes of automation in terms of cleared aircraft, aircraft conflicts, and aircraft collisions. This finding was counter to our expectation (H1). However, it is possible that a disproportionate number of manual and automated control periods during the AA trials lead to AA performance (on average) to approximate manual control. The AA trials involving information acquisition, information analysis, and decision making aiding all produced more manual control minutes than automated control minutes (59%, 63%, and 60% manual minutes, respectively). It is possible that the large numbers of manual control minutes as part of the AA conditions may have caused any differences between the modes of automation and the control condition to be indiscernible.

Results of ANOVAs on data collected during the automated control periods as part of the AA conditions (only) revealed a significant effect of mode of automation on the number of cleared aircraft ($F(3,53)=4.03$, $p=0.0118$) and the number of aircraft collisions ($F(3,51)=3.02$, $p=0.0382$). An ANOVA on the number of aircraft conflicts revealed no significant effect of the mode of automation. Figure 11 summarizes the mean number of aircraft cleared, conflicting, and colliding under each mode of AA during the automation control periods of the Multitask© simulation.

It is apparent from the plot that, as hypothesized, the action implementation mode of automation, a lower-order sensory/response function, yielded higher average performance during automated control periods in terms of cleared aircraft as a result of the automated “hand-off” of aircraft control from TRACON to local tower control. A Tukey’s test confirmed that action implementation was significantly better ($p<0.05$) in terms of cleared aircraft, as compared to the information analysis mode of automation, a higher-order air traffic controller cognitive function (H2). Tukey’s test also revealed information analysis to be significantly inferior ($p<0.05$) for preventing collisions. This finding further supported the hypothesis that AA would be more effective when applied to lower-order sensory/response functions (H2). These findings are consistent with the results of Leroux (1993), Laois and Giannacourou (1995), Kaber, Prinzel et al. (2002), and Clamann and Kaber (2003).

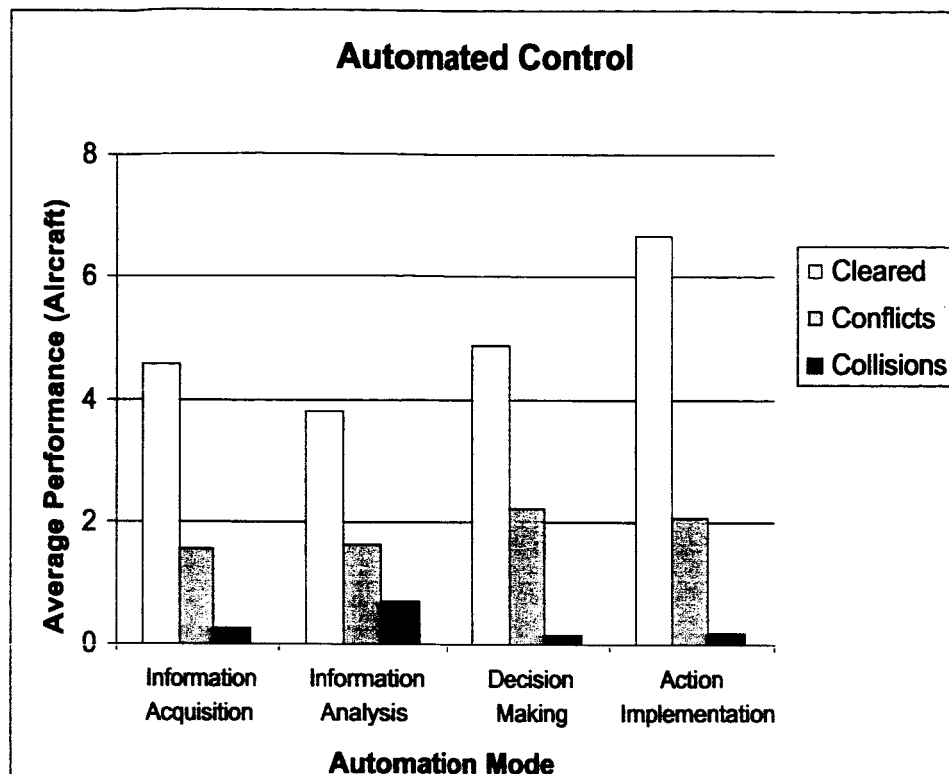


Figure 11. Primary task performance during automated control periods

Residual diagnostics on the cleared aircraft data collected during manual control periods as part of the AA trials revealed non-constant variance across the modes of automation. A square root transformation was performed on the response data (Neter, Wasserman, & Kutner, 1990, p.146) in order to account for the potential ANOVA assumption violation. Analysis of Variance results revealed a significant effect of mode of automation ($F(3,53)=3.73$, $p=0.0166$) on the square root of cleared aircraft. Mode of automation did not prove to have a significant effect on the number of aircraft conflicts or collisions during manual control periods. Figure 12 shows the mean values for the number of aircraft cleared, conflicting, and colliding during manual control periods of the AA simulation trials.

Contrary to hypothesis (H2), the plot shows greater average performance during trials involving the higher-level information analysis mode of AA in terms of cleared aircraft. A Tukey's test confirmed that subjects were significantly better ($p<0.05$) at clearing aircraft during manual control periods, when AA was applied to the information analysis aspect of the ATC task, versus AA of the action implementation function. There appeared to be a positive carry-over effect of information analysis automation on subject ability to clear aircraft during manual control periods as part of AA. However, this was not the case for subject ability to deal with negative events including conflicts and collisions. On average, there was a greater number of conflicts and collisions during the manual periods of the information analysis condition, but any differences among the AA conditions in terms of these variables were not supported by statistical evidence. It is possible that these findings could be attributed to the fact that the average duration of manual control periods when AA was applied to the information analysis functions was longer

than any other AA condition. The lack of additional significant findings on the manual control periods may be attributed to the nature of manual control remaining consistent across the various modes of AA. Clamann and Kaber (2003) also observed no significant findings relating to manual control periods of AA.

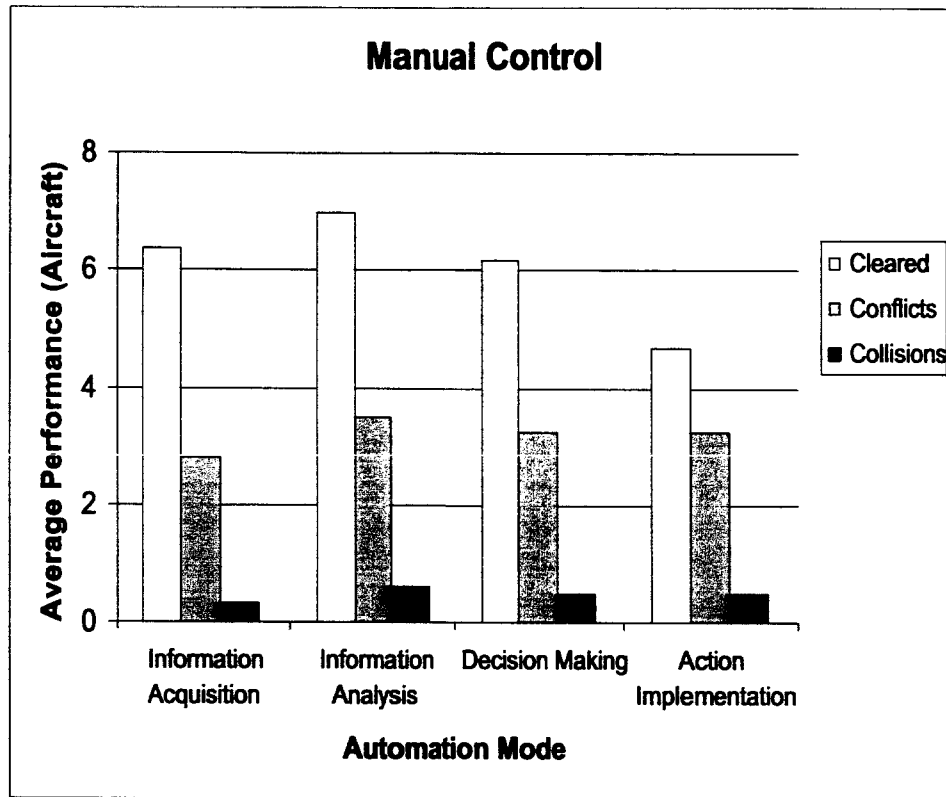


Figure 12. Primary task performance during manual control periods.

Workload (Secondary-task Performance)

An ANOVA on the workload data did not reveal significant effects due to the mode of automation when comparing the manual control condition with the AA conditions or when analyzing the automated and manual control periods as part of AA. These results are surprising as the modes of automation were expected to decrease workload (increase average secondary-task performance) compared to the manual control condition (H3). Likewise, the information acquisition and action implementation modes of automation were expected to decrease workload compared to the other modes of automation (H4, H5).

Hilburn et al. (1997) found experiment trials involving manual control to produce higher workload than AA conditions. Kaber, Prinzel et al. (2002) observed higher levels of workload when AA was applied to the information analysis function of the primary task compared to the information acquisition and action implementation modes of automation. Anecdotal observations during the present experiment suggested high workload associated with the information analysis mode of automation due to the large amount of data displayed during these trials. Related to this, information acquisition was identified as the preferred mode of automation by most subjects.

However, there was no statistical evidence to support these observations.

This lack of significant findings may be attributed to varying strategies employed by subjects to monitor the secondary task. There were significant individual differences in secondary-task performance revealed through the comparison of the completely manual condition with AA conditions ($F(7,68)=16.97$, $p<0.0001$), and in analyzing the automated control periods ($F(7,52)=8.19$, $p<0.0001$) and manual control minutes ($F(7,68)=11.89$, $p<0.0001$) separately.

Situation Awareness

With respect to SA under manual versus automated control periods, an ANOVA revealed a significant effect of the general type of control on the arcsine of percent correct responses to Level 3 SA queries ($F(1,216)=9.33$, $p=0.0025$); however, there were no significant effects on the percent correct responses to Level 1 and 2 SA queries or for total SAGAT scores.

Figure 13 summarizes the mean Level 1, Level 2, Level 3 and total SA scores for both automated and manual control periods. As hypothesized (H6), the plot reveals that subjects were, on average, better at answering Level 3 SAGAT queries during manual control periods compared to automated control periods. This finding supports the notion that introducing automation in ATC may remove the controller from the control loop (Endsley & Kaber, 1999) and lead to decrements in a controller's SA. However, results on Level 1 and 2 SA were not supportive of our general hypothesis.

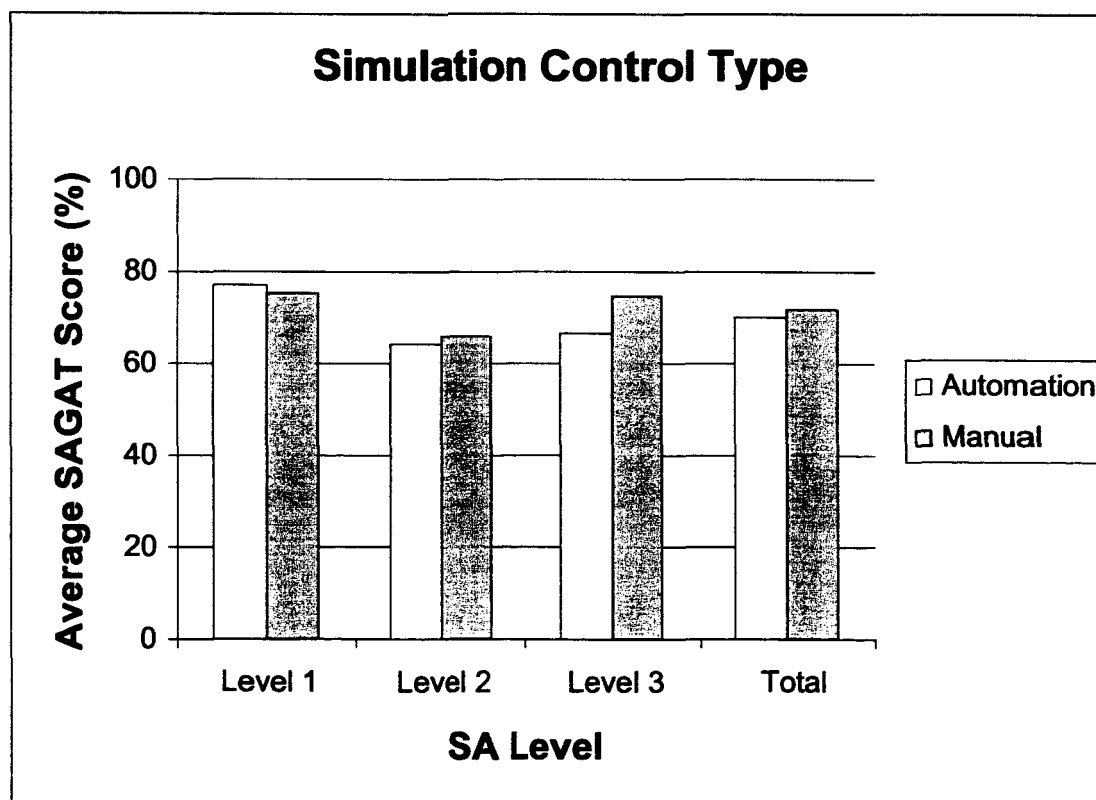


Figure 13. Mean SAGAT scores during manual and automation control periods.

Surprisingly, none of the aggregate SA response measures revealed a significant effect of the specific forms of AA (H7). This was unexpected because the classification of various queries according to Endsley's (1995a) levels of SA was considered to be accurate based on the type of dynamic knowledge questions called for.

The SAGAT data was subsequently analyzed on a question-by-question basis for effects of the various forms of AA, resulting in 18 separate analyses. Most surprisingly, ANOVA results revealed no significant effect of mode of automation on the arcsine transform of percent correct responses for all queries. There were, however, significant individual differences observed in responses to the majority of SA queries. It was expected that SA would decrease during action implementation, as the controller may have been OOTL compared to other modes of automation (H8). In addition, Level 2 SA was expected to be worse under the information analysis LOA (H9), and decision making automation trials were expected to decrease the accuracy of responses to Level 3 SA questions (H10). However, neither of these expectations was supported by the results on the SA data.

This was unexpected because previous research on SA in ATC has demonstrated SAGAT to be a valid and sensitive indicator of complex system operator perception, comprehension, and projection for varying workload and display conditions in other domains (Endsley & Kiris, 1995; Jones & Endsley, 2000). However, some more recent work by Hauss and Eyferth (2003) suggests that SAGAT may not be a sensitive measure for SA in the ATC environment due to different aircraft having different relevance to controllers at different times in a simulation.

Hauss and Eyferth (2003) suggested that controllers may use an event based mental representation of the air traffic situation in order to determine what information is currently relevant, what information will be relevant in the future, and what information can be neglected in an attempt to make the task manageable from a working memory perspective. In addition, Hauss and Eyferth (2003) state that aircraft which have recently been contacted by a controller and that have required recent control actions, or aircraft that are currently (or will soon be) in conflict, may demand more attentional resources than other display aircraft. Consequently, controllers may focus on certain aircraft to the exclusion of others at various times during control activities. Hauss and Eyferth (2003) resolved that controllers will recall the flight parameters of aircraft most relevant to current task performance more accurately in responding to SAGAT queries than the parameters for aircraft that are not critical to ATC at the time of the SAGAT freeze. Gronlund, Ohrt, Dougherty, Perry, and Manning (1998) said that, "not all aircraft are equally important to the controller, and measures of SA should not assume that they are." It is possible that in the current study subjects responded accurately to SAGAT queries for the aircraft that they had recently dealt with and poorly for those aircraft not in the focus of their attention. Since each SAGAT query was posed for all 7 aircraft on the display at a freeze, with respect to our data analysis, the average percent correct responses to a single query could have remained relatively constant across the experimental conditions as a result of this aircraft relevance issue.

Beyond this, Hauss and Eyferth (2003) state that expert controllers actually remember less about the ATC environment compared to novice controllers, because expert controllers, based on experience, know what aircraft are relevant and demand a heightened amount of attention.

Niessen, Eyferth and Bierwagen (1999) agree that experience plays a large roll in controller SA, as experienced operators “use less information, but that this information is diagnostically more relevant” (p. 1513). The subjects in this study were not expert controllers, but they did receive 3.5 hours of training on the Multitask© simulation. Based on the residual and trial number plots, no learning was found, suggesting that the subjects did achieve some level of expertise in the ATC simulation. Therefore, the subjects may have remembered only what information was currently relevant to their clearance amendments. Using a SA measure that placed equal emphasis on aircraft they considered irrelevant at certain points in task performance may have been an insensitive measurement approach for the domain. Based on similar concerns, Hauss and Eyferth (2003) actually presented a method for SA assessment, which weights aircraft based on their relevance to the current control scenario.

4. Experiment #2

Based on the results of the first experiment, we conducted a second experiment using a modified approach to the SAGAT-based measurement of operator SA in the ATC simulation. Unfortunately, the first experiment did not provide substantial evidence of the sensitivity of the SAGAT measure (without weighting of the relevance of aircraft for controllers at the times of simulation freezes). We wanted to devise a measurement approach that would allow us to effectively and reliably identify any SA implications of manual control and AA, in general. We also needed to refine the measurement approach to effectively quantify fluctuations in SA in the ATC task resulting from DFAs (associated with operator workload changes) applying automation to the various Multitask© information processing functions.

The key problems that we identified with the SA measurement approach as part of the first experiment included:

- (1) The technique assumed that the aircraft subjects recalled in labeling the graphic of the Multitask© radarscope were those aircraft they considered to be most relevant to their goals at the time of the simulation freeze.
- (2) SA queries were posed for every aircraft that was recalled by a subject and identified on the graphic of the radarscope. If a subject remembered and correctly identified the locations of all aircraft, they were asked questions on all of them. However, if subjects could only recall a subset of the aircraft currently being controlled, they only answered SA queries on those aircraft whose locations were correctly recalled.
- (3) Subjects were not penalized for aircraft that they could not recall or draw accurately on the radarscope graphic (i.e., the percent correct responses were calculated as the number of correct answers divided by the actual number of queries posed (not by the number of possible queries, which would have taken into account all aircraft on the screen)).

We suspected that the method of subject recall of aircraft and the implementation of the SAGAT measure as part of the first experiment might have contributed to the limited number of SA findings.

Consequently, we conducted an additional literature search for alternative measurement approaches. We found that other recent work had investigated the use of SAGAT for ATC tasks.

Nunes (2003) assessed the performance and SA of ATC trainees in an experimental task in which they were either aided or unaided when evaluating pilot requests for flight plan deviations. Trainees in the “aided” condition were given a datalink trajectory of the proposed deviation, while those in the “unaided” condition had to extrapolate the trajectories. Additionally, the aided and unaided conditions were tested at both high and low levels of workload. Surprisingly, Nunes (2003) found that SA, as measured by the SAGAT technique, did not vary between display conditions, while the performance measures of response time and response accuracy revealed a significant effect of display condition.

Nunes (2003) argued that SA may not have varied between groups because both aided and unaided controllers essentially processed the same information, but it was the level at which the processing occurred that ultimately affected the accuracy scores. Therefore, unaided controllers processed information at a greater level of depth, resulting in a higher performance score being obtained. This research suggested that the nature of the interface may compromise or enhance the controller’s mental model, even though SA is not affected. However, Nunes (2003) did not provide an explanation of how “information processing” could be affected by the aiding condition without any effect on SA, as operationally defined through the SAGAT method. It is also plausible that the results Nunes (2003) obtained could be attributed to the SAGAT measure not being sensitive enough for use in this domain or task structure.

As previously mentioned, Hauss and Eyferth (2003) demonstrated that there may be more suitable measures of SA in the ATC domain than the approach we took in our first experiment. Hauss and Eyferth (2003) contended that SAGAT is based on the assumption that the set of task environment elements relevant for SA is independent of the dynamics within the task environment. They developed a new SA measure for ATC called SALSA, and compared SAGAT and SALSA in an air traffic management (ATM) study. SALSA differed from the SAGAT approach in two ways. First, it involved an expert rating of replay of the ATM simulation to determine the relevance of each task element (aircraft) to controllers. Only elements that were judged as relevant in the replay were considered for SA queries. Second, rather than having subjects recall aircraft positions on a blank radarscope, SALSA involved cued recall, in which the subjects were given the positions for the aircraft they were to be queried on. Hauss & Eyferth (2003) contended that these changes reduced the possibility of subjects confusing two aircraft positions during free recall and took into account the air traffic controllers use of an event-based mental representation of their task. Their empirical results confirmed that controllers used event-based representation, since significantly more relevant parameters than irrelevant parameters were reproduced using the SALSA measure. Further, the SALSA measure responded to changes in workload in intuitive, although not statistically significant, ways.

On the basis of this research, we decided to investigate an alternate approach to implementation of our SAGAT-based approach for SA assessment in the Multitask© simulation. In general, the modifications to the measurement method included: (1) cueing of aircraft positions, in comparison to free recall; and (2) objective weighting of relevance for those aircraft to be queried. While the Hauss & Eyferth (2003) study used expert ratings of a playback of the simulation to determine relevance, the approach taken in this experiment involved designation of aircraft currently in conflict, as well as those that had recently been issued clearances (e.g. holding fixes, speed reductions, and runway changes) in the simulation as a real-time measure of

relevance. It was expected that these modifications would lead to a more sensitive assessment of the impact of AA on controller SA.

Participants

As in the first experiment, we recruited a small sample of 8 subjects for participation. Subjects ranged in age from 21 to 29 years. All subjects had 20/20, or corrected to normal, vision and personal computer experience. None of the subjects had flying or ATC experience. Subjects were compensated at a rate of \$7.5 per hour for their participation (due to the limitations of research support funds).

Tasks and Equipment

The tasks used in this experiment were identical to the tasks used in Experiment #1, except for the following modifications to the Multitask© simulation:

- (1) A predefined control sector was presented as a hexagon overlay on the radarscope near the 50 and 60 nm bands. All aircraft initially appeared at the intersection of their randomly assigned approach trajectory and the sector envelope. This substantially reduced the total area of control for subjects and accelerated the action as part of the scenario.
- (2) A 25 s delay was introduced between the appearances of each new aircraft on the display in order to prevent aircraft conflict situations at the periphery of the control sector.
- (3) Under the information analysis and decision making modes of automation, auditory alerts of potential aircraft collisions were provided to operators. This served to make conflicts more salient to subjects, when they may have been uncertain about a potential collision of two aircraft, based on the predefined conflict criteria (e.g., 3nm of lateral separation). (The use of these cues did not go against the theoretical descriptions of the decision making and information analysis forms of automation provided by Parasuraman et al. (2000).)

The experiment environment and equipment setup used in this experiment was also identical to that used in Experiment #1. The primary workstation was integrated with the digital flat panel monitors and two sets of interface controls (keyboards and mice). The secondary computer station was once again used for the SA query administration. (The specific steps as part of the implementation of the SA measure are discussed below.)

Approach to AA

The gauge-monitoring task was run concurrently with the Multitask© simulation, and used as an objective measure of workload to trigger DFAs in the ATC task (i.e., automation was activated when workload was high, and manual control was reinstated when workload was low). The gauge performance criteria used in this experiment for driving AA were the same as those used in Experiment #1.

As in Experiment #1, the final dual-task practice session (on the first day of the experiment) was used to establish the average gauge-monitoring performance and the SD for pairs of subjects under randomly assigned modes of automation. Their performance was used, in turn, to establish Multitask© “overload” and “underload” conditions for all subjects exposed to the same mode of automation during test trials.

Design of Experiment and Variables

Like Experiment #1, the current experiment followed a completely within-subjects design with blocking on the subject. All eight subjects completed two, 30-min. trials under each of the five modes of Multitask© control. The IV manipulated in this experiment was identical to the IV in Experiment #1. Adaptive automation was applied to Multitask© simulation functions, including information acquisition, information analysis, decision making, and action implementation. These conditions were compared with a completely manual control condition.

The dependent variables observed during this experiment were also identical to those measured in Experiment #1, except for the SA measures. All Multitask© and gauge-task performance measures were recorded on a per minute basis. With respect to the SA measures, we used a modified version of the SAGAT. As in Experiment #1, simulation freezes were conducted at random points in time during experimental trials in order to deliver SA queries to subjects. However, in this experiment, subjects were posed with 9 questions during each freeze, including 3 queries targeting each level of SA (1 – perception; 2 – comprehension; 3 – projection). When a freeze occurred, subjects were asked to move to the secondary computer workstation and use the database application to respond to queries. At the same time, an experimenter collected information from the Multitask© software by accessing a CDA/CAA aid, which was hidden from subjects during testing. The aid provided information on aircraft in conflict with each other and recommended clearances. Based on this information, the experimenter identified the three aircraft with the highest priority, or greatest “relevance”, at that point in time in the simulation. The following detailed hierarchy of simulation events was used as a basis for determining the relevance of aircraft:

- (1) Aircraft that were currently in conflict, and those that had been in conflict the longest, were considered to have the highest priority for subjects.
- (2) Aircraft that had been issued a “hold” clearance were considered to have the next highest priority. This amendment impacted the flight path as well as landing time, and subjects needed to issue an additional clearance to land the aircraft.
- (3) Aircraft that had been issued a “change airport” clearance were considered to have the next highest priority. Similar to the “hold” clearance, this amendment affected the flight path and landing time.
- (4) Aircraft that had been issued a “reduce speed” clearance were considered to have the next highest priority. Reducing an aircraft’s speed had immediate and salient consequences for a controller.
- (5) Aircraft that had been issued a “change runway” clearance were considered to have the next highest priority. This amendment usually did not represent an immediate emergency, as did “reduce speed”. Aircraft collisions on the runway were also less common than other collisions.

- (6) The last aircraft queried by the subject was considered to have the lowest priority. Subjects often queried aircraft to which they were considering issuing a clearance.

This approach to cueing and prioritizing aircraft for queries was similar to that used by Hauss and Eyferth (2003). For consistency, subjects were always queried about three aircraft at each freeze. If less than three aircraft matched the aforementioned criteria, random aircraft were selected for querying; however, subjects' answers to these queries were not considered in the final SA analysis, as the relevance of these aircraft to the subject was not known.

To facilitate the cued recall of aircraft as a basis for questioning, at each freeze an experimenter quickly sketched the locations of the "high priority" aircraft on a blank drawing of the Multitask© radarscope. The subjects were then given the drawing of the radarscope and asked to respond to each of the 9 SA queries for each "high priority" aircraft.

During the test trials, the experimenter also recorded information on each aircraft, including speed, destination airport, destination runway, and any clearance amendments. This information was used for selecting aircraft for querying and to later evaluate subjects' answers to the queries.

Procedures

The steps in the procedure of this experiment were identical to those of Experiment #1. Multitask© and gauge-monitoring training were conducted, including subject practice under each of the four modes of Multitask© automation, completely manual control and AA. Beyond this, subjects completed a dual-task practice trial, in which they performed the Multitask© manually and under a randomly assigned mode of automation. Just before the second dual-task practice session, subjects were familiarized with the SAGAT, and sample questions were presented. The subjects then experienced three SAGAT freezes during the practice session. As in Experiment #1, the SAGAT questions presented during the practice session were randomly selected from the questions to be posed during the actual experimental trials. As previously mentioned, this session also yielded the gauge performance criteria for triggering AA during test trials.

During the second and third days of the experiment, subjects reviewed the Multitask© simulation procedures and completed 10 experimental trials involving the use of the various modes of Multitask© automation. The order in which the modes were presented to subjects was randomized. Each trial lasted for approximately 50-60 minutes, including 30 minutes of simulation time and approximately 20-30 minutes to answer SAGAT queries during freezes. Again, the Multitask© skill level was set to seven (7 aircraft) for all training, practice, and experiment sessions.

Like Experiment #1, this study lasted approximately 3 weeks. At the conclusion of the last session, each subject signed their payment form and they were debriefed.

Data Analysis

All performance, workload and SA response measures were summarized for completely manual control trials and automated and manual control periods as part of AA trials. The entire

experiment yielded 80 performance observations across all subjects (8 subjects x 5 modes of control x 2 trials) and 64 automated performance scores resulted from the AA conditions (8 subjects x 2 trials x 4 automation modes). With respect to SA, during each trial subjects answered 81 SAGAT queries (3 freezes x 9 queries x 3 "high-priority" aircraft).

As in the data analysis as part of Experiment #1, ANOVAs were conducted on all response measures, including the SAGAT data, with the mode of AA and subject included as predictor variables in the ANOVA model. In order to demonstrate sensitivity and reliability of the modified SA metric, the SAGAT data from the experiment was used to establish whether operator accuracy in responding to queries was predictable based on the allocations of manual or automated control as part of the AA conditions. Duncan's Multiple Range (MR) tests were used to breakout the means associated with the various AA conditions and manual control condition. Finally, correlation analyses were conducted on the performance and SAGAT response measures. In addition to the ANOVAs, we considered correlation coefficients as a vehicle for assessing the sensitivity of the SAGAT-based measure. We wanted to determine whether changes in operator SA under AA corresponded with any changes in the number of aircraft conflicts and clearances, etc. across manual and automated control periods in the Multitask© simulation.

Diagnostics and statistical model

As in Experiment #1, we used SAS to generate ANOVA model residuals and predicted values. Residual and normal probability plots were used to investigate any potential ANOVA assumption violations. These diagnostics indicated if any transformations on the response or IV might be required to ensure the appropriateness of the parametric tests, particularly for the SAGAT data. Any specific transformations employed in the various analyses are described in the Results section. As mentioned in the Data Analysis section for Experiment #1, since the responses to SAGAT questions represent a binomial variable (correct or incorrect), the discrete nature of the data violates the ANOVA assumptions. Per Endsley (1995b), we applied an arcsine transform to the percent correct responses for each query. Finally, based on plots of residuals against trial number, no learning effects were found in either the Multitask© or gauge-monitoring performance data sets.

Specific Hypotheses

The hypotheses for this experiment were similar to those posed for Experiment #1 and we list them here:

E2H1 - Although the results of Experiment #1 did not reveal AA to be superior to manual control, we suspected the lack of a significant difference may have been attributable to individual differences. On the basis of prior research, we expected AA to yield a greater number of successfully cleared aircraft and fewer conflicts and collisions than manual control.

E2H2 - Adaptive automation of the Multitask© was expected to affect performance on the secondary gauge-monitoring task, or operator workload. Although Experiment #1 did not

reveal significant differences among manual control and automation, again we suspected that the lack of a significant difference may have been attributable to substantial individual differences in strategies to performing the Multitask© simulation under particular modes of automation. In general, we expected manual control to result in higher levels of operator workload than the AA trials.

E2H3 – We expected the modified approach to SAGAT in the Multitask© simulation to be sensitive to changes in controller SA as a result of the AA manipulations. In general, subjects were expected to do better at responding to SA queries under lower levels of automation (information acquisition) and manual control as compared to high-level automation (information analysis and decision making) because of the potential for OOTL performance problems. Information acquisition automation was designed to maintain controller involvement in the system control loop and it presented the TPA, which was expected to draw operator visual attention to the display and promote concentration. We remained skeptical of whether the action implementation mode of automation would improve controller SA. Again, since this mode automatically “handed-off” aircraft to tower control, eliminating the risk of a collision within 20 nm of the radarscope center, it removed the controller from the control loop and had the potential to adversely affect SA.

E2H4 – We also speculated that under high levels of automation, such as decision making or information analysis, operators would exploit the additional capabilities of the automation, including following computer assisted recommendations on clearances for specific aircraft. With this in mind, subjects were expected to pay less attention to the actual radarscope and focus on the decision aid display. The same was expected for the information analysis condition, which presented operators with a display of summary data on simulated aircraft as well as information on the criticality of issuing clearances to particular aircraft. Our hypothesis was that these forms of Multitask© automation would remove operators from the low-level control functions, including watching the scope and querying aircraft, which may be important to achieving SA, as measured using the queries.

Results and Discussion

Primary Task Performance

An ANOVA on Multitask© performance revealed no significant differences between the completely manual control condition and the modes of automation in terms of cleared aircraft, aircraft conflicts, and aircraft collisions. This finding was counter to our expectation (E2H1). As in Experiment #1, it is possible that a disproportionate number of manual and automated control periods during the AA trials lead to AA performance (on average) approximating manual control. The AA trials involving information acquisition, information analysis, decision making, and action implementation all produced more manual control minutes than automated control minutes (81%, 86%, 65%, and 72% manual minutes, respectively). It is possible that the large numbers of manual control minutes as part of the AA conditions may have caused any differences between the modes of automation and the control condition to be indiscernible. It is also likely that averaging low performance scores under automated control periods with high scores under manual control periods in AA trials, and vice versa, washed-out any significant

differences among the conditions in this total performance analysis (see separate results on manual and automated control periods below).

Results of ANOVAs on data collected during the automated control periods as part of the AA conditions revealed a significant effect of mode of automation on the number of cleared aircraft ($F(3,41)=3.62$, $p=0.0208$) and the number of aircraft conflicts ($F(3,41)=3.97$, $p=0.0143$). An ANOVA on the number of aircraft collisions did not reveal a significant effect of mode of automation. Figure 14 summarizes the mean number of aircraft cleared, conflicting, and colliding under each mode of AA during the automation control periods of the Multitask© simulation.

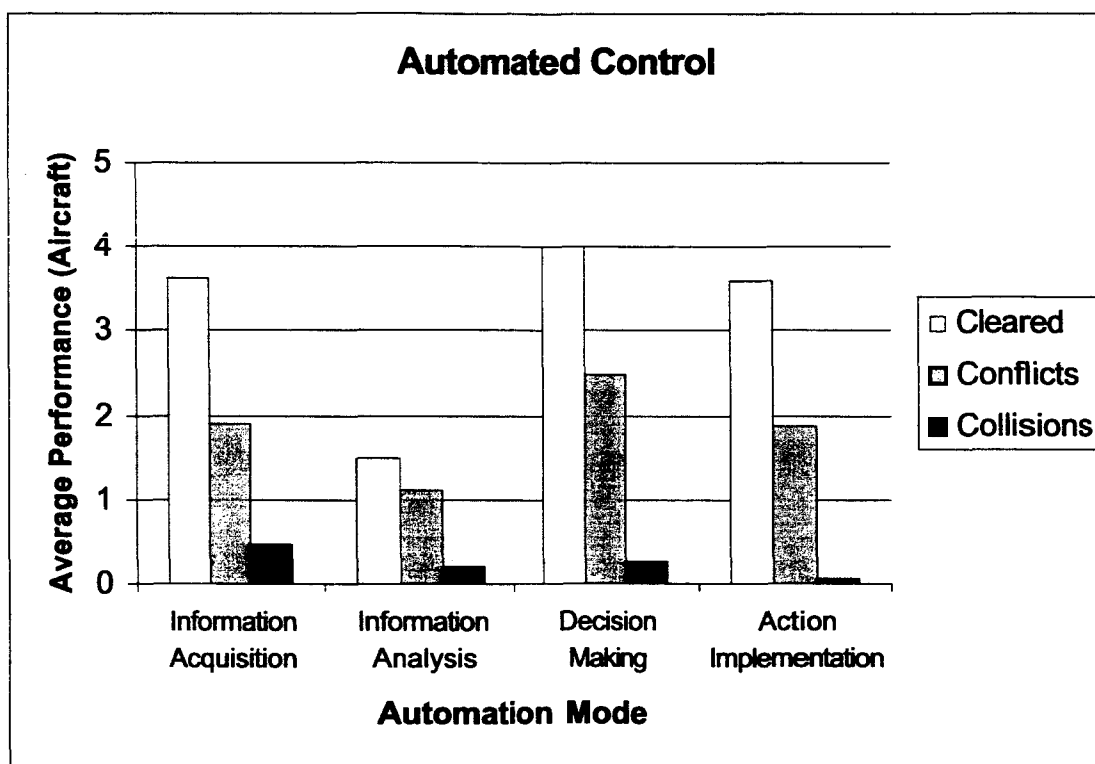


Figure 14. Primary task performance during automated control periods

The plot shows the number of cleared aircraft to be higher for the information acquisition, decision making, and action implementation modes of automation. Duncan's MR test confirmed that these modes of automation were significantly superior ($p<0.05$) in terms of cleared aircraft, as compared to the information analysis mode of automation. Use of the information analysis mode, during automation periods, also produced the worst performance in Experiment #1. These results support our hypothesis from that experiment, which stated that superior Multitask© performance was expected during trials in which AA was applied to information acquisition and action implementation (H2), the lower-order sensory/response functions. The high number of cleared aircraft during decision making may be attributable to the longer automated control periods under this mode of automation, as compared with the other modes.

Duncan's test also revealed decision making to be significantly worse than information analysis for preventing aircraft conflicts ($p < 0.05$). This finding was not surprising given that the decision aid made recommendations to subjects for dealing with conflicts once the computer detected them. It is possible that subjects developed a strategy of waiting for the automation to warn them of a conflict and then to think about how to appropriately clear aircraft. This is, of course, counter to the practice of actual controller's and would've been counter to our experiment instructions. The small number of conflicts during information analysis automation may be the result of shorter automated control periods under this mode, as compared with the other modes of automation. That is, there was simply less time for conflicts to develop and be recorded under this mode.

Analysis of Variance results on manual control periods as part of the AA conditions revealed a significant effect of mode of automation on the number of cleared aircraft ($F(4,68)=7.58$, $p < 0.0001$). As in Experiment #1, an ANOVA on the number of aircraft conflicts and collisions revealed no significant effect of the mode of automation. That is, subjects appeared to be relatively consistent in managing negative events in the simulation across the manual periods as part of the AA trials. Figure 15 summarizes the mean number of aircraft cleared, conflicting, and colliding under each mode of AA during the manual control periods of the Multitask© simulation. This plot is nearly identical to the plot of the same type of data for Experiment #1, save the mean observations on the decision making condition.

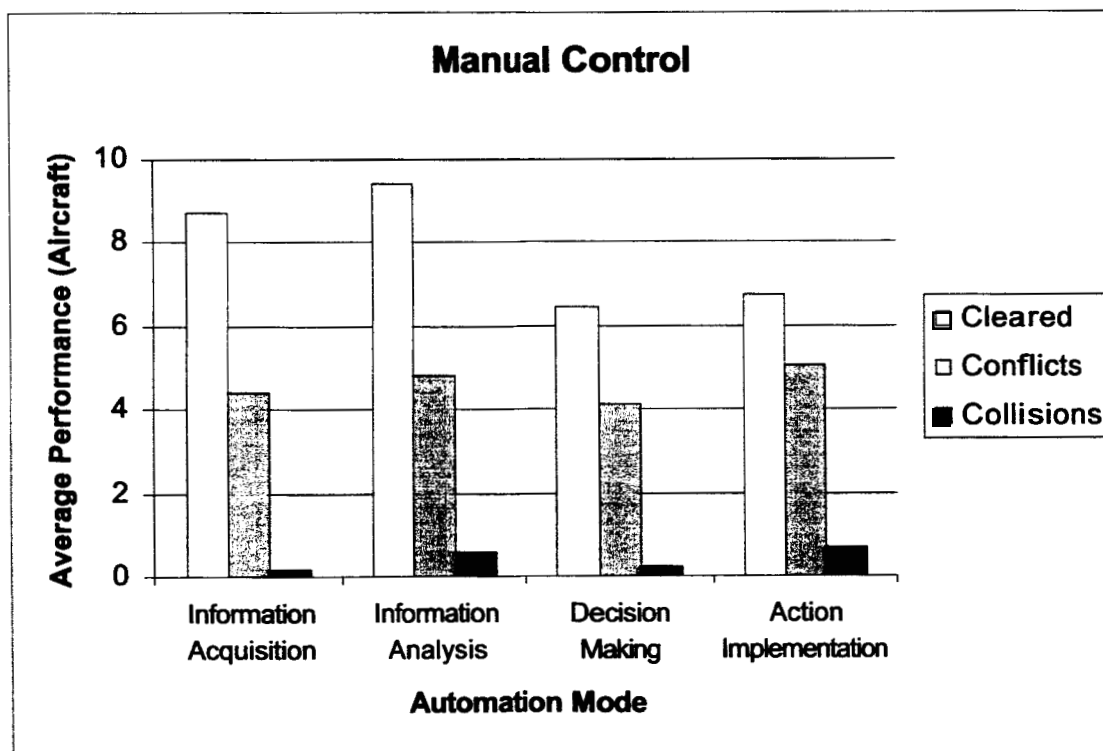


Figure 15. Primary task performance during manual control periods

The number of cleared aircraft was significantly higher for the information acquisition and information analysis modes of automation than for decision making and action implementation

(Duncan's test, $p < 0.05$). The results on the decision making condition are in agreement with our original hypothesis that AA applied to lower-order information processing functions would yield better performance. It is possible that there was a carry-over effect of the decision making automation on the manual control periods. Subjects may have needed time to re-orient to the manual mode, when the decision aid disappeared, and they had to identify conflicts themselves. This change in effort (or the subject's role) may have subtracted from the time they allocated to actually clearing aircraft and, consequently, degraded performance along that dimension. At a more cursory level of inference, the low number of cleared aircraft under action implementation, as well as the high number of cleared aircraft under information analysis, may be explained by the relative amount of time subjects spent in manual control as part of AA during these modes of automation (72% manual control minutes under action implementation compared with 86% manual control minutes under information analysis). There was simply less time spent under manual control, when AA was applied to the action implementation function, and consequently there were fewer observations of clearances for these periods. It is also possible that there was a carry-over effect of the action implementation automation on the subsequent manual control periods as part of AA trials. Subjects may have become accustomed to the automated "hand-off" of aircraft to tower control, as part of the action implementation automation, once a vehicle passed within 20 nm miles of an airport. Subjects might have forgotten that this type of control was not active under the manual mode, or they had to re-orient to issuing additional types of clearances to aircraft (e.g., change of runway) when vehicles were close to airports. This may have compromised the number of successful clearances.

The ANOVAs on the data collected during the manual and automated control periods did reveal significant effects of subject on the number of cleared aircraft (manual: $F(7,68)=6.83$, $p < 0.0001$; automated: $F(7,41)=4.76$, $p = 0.0006$). With this in mind, in order to further control for individual differences in strategies for performing the Multitask under the various modes of auto (e.g., some subjects sending many more aircraft to holding fixes than others), the ratios of cleared aircraft, conflicts, and collisions to the total number of aircraft presented during a trial were calculated and the ANOVAs were applied to these response measures with LOA & subject included as IVs in the statistical model. The pattern of results on the ratio measures was similar to that observed on the original performance measures for manual, automated, and total Multitask performance.

Workload (Secondary-task Performance)

An ANOVA on the workload data combined across manual and automated control periods did not reveal significant effects due to the mode of automation when comparing the manual control condition with the AA conditions (E2H2). (It is important to note here that 12 observations were missing from the gauge-monitoring task data for the automated control periods, as 12 automation trials did not contain any automated minutes. This decreased the number of automated performance scores for the AA conditions from 64 to 52.) In general, we expected manual trials to result in higher levels of operator workload than the AA trials. As in Experiment #1, the lack of a significant difference when looking at the combined data may have been attributable to significant individual differences ($F(7,68)=3.16$, $p = 0.0059$) in strategies of performing the Multitask© simulation. However, it is more likely that simply averaging low gauge scores with high scores, under high-level automation (information analysis and decision making), for automated and manual control periods, respectively, as well as averaging high gauge scores with

low scores, under low-level automation (information acquisition and action implementation), for automated and manual control periods, respectively, washed-out any significant differences among the conditions in this total performance analysis (see separate results on manual and automated control periods below).

An ANOVA on the workload data did reveal significant effects due to the mode of automation when analyzing the automated control periods as part of AA ($F(3,41)=4.01$, $p=0.0137$). Figure 16 summarizes the mean hit-to-signal ratio under each mode of AA of the Multitask© simulation.

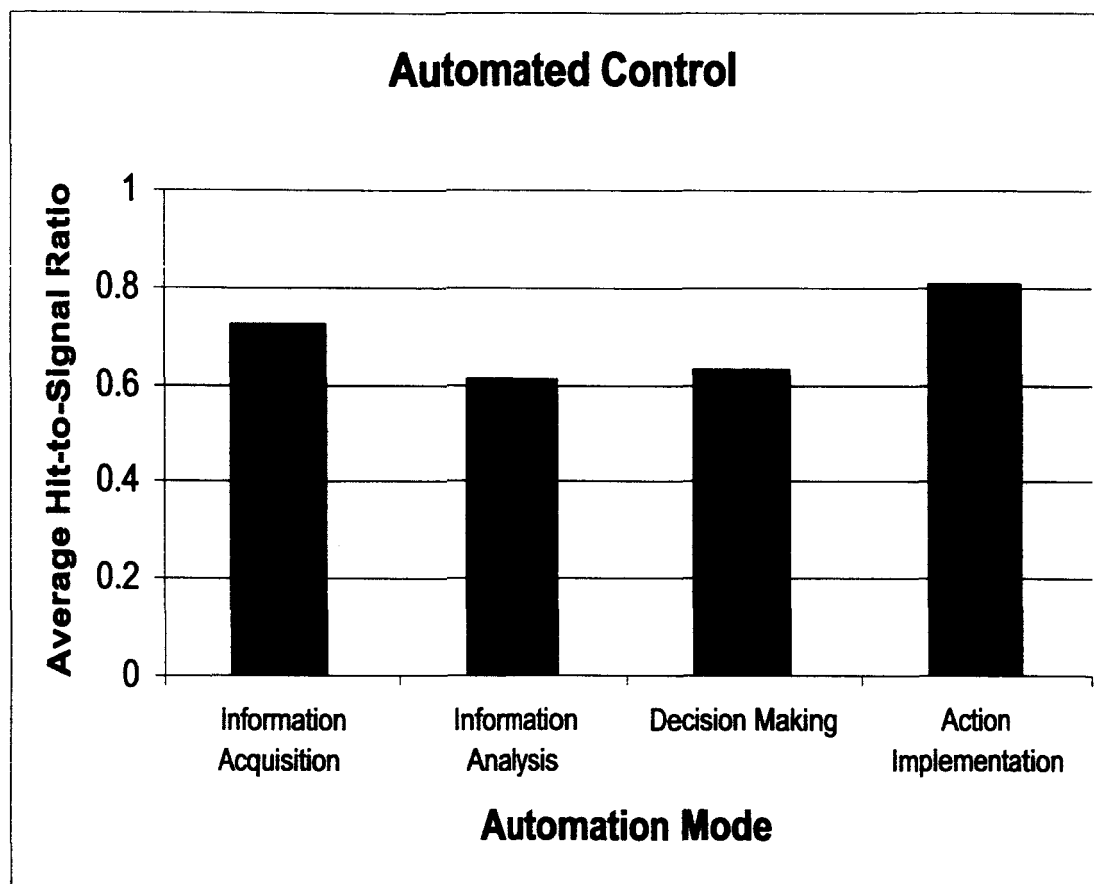


Figure 16. Secondary-task performance during automated control periods.

It is apparent from the plot that the action implementation mode, a lower-order sensory/response function, yielded higher average secondary-task performance during automated control periods, suggesting that the workload during these periods under this mode was lower than the workload in the automated periods under other modes. Duncan's MR tests confirmed that secondary-task performance was significantly higher under action implementation than in information analysis and decision-making modes, which were considered higher-order information processing functions. These findings are in line with hypotheses posed for Experiment 1. Based on Kaber, Prinzel et al. (2002) results, and observations made by Parasuraman et al. (2000), it was expected

that higher levels of automation, including information analysis and decision making, presenting complex displays for operator interpretation might demand high levels of visual attention and actually increase workload (as evidenced by decreases in gauge-monitoring task performance) (H4). Related to this, we expected modes of automation not presenting large amounts of information for operators to process (e.g., action implementation), or presenting information directly on the radar display (e.g., information acquisition), to decrease workload (H5).

An ANOVA on the workload data also revealed significant effects due to the mode of automation when comparing the manual control condition with the manual control periods as part of AA ($F(4,68)=2.66$, $p=0.0399$). Residuals plots revealed one potential outlier among the 52 observations as well as a potential violation of the normality assumption of the ANOVA, as indicated by a significant Shapiro-Wilk's test ($p=0.0086$). However, the extreme data point was retained in the data set because there was no experimental anomaly as a basis for removal, and the potential assumption violation was not deemed severe enough for transformation of the entire data set. Figure 17 summarizes the mean hit-to-signal ratio under the manual control periods of each test trial.

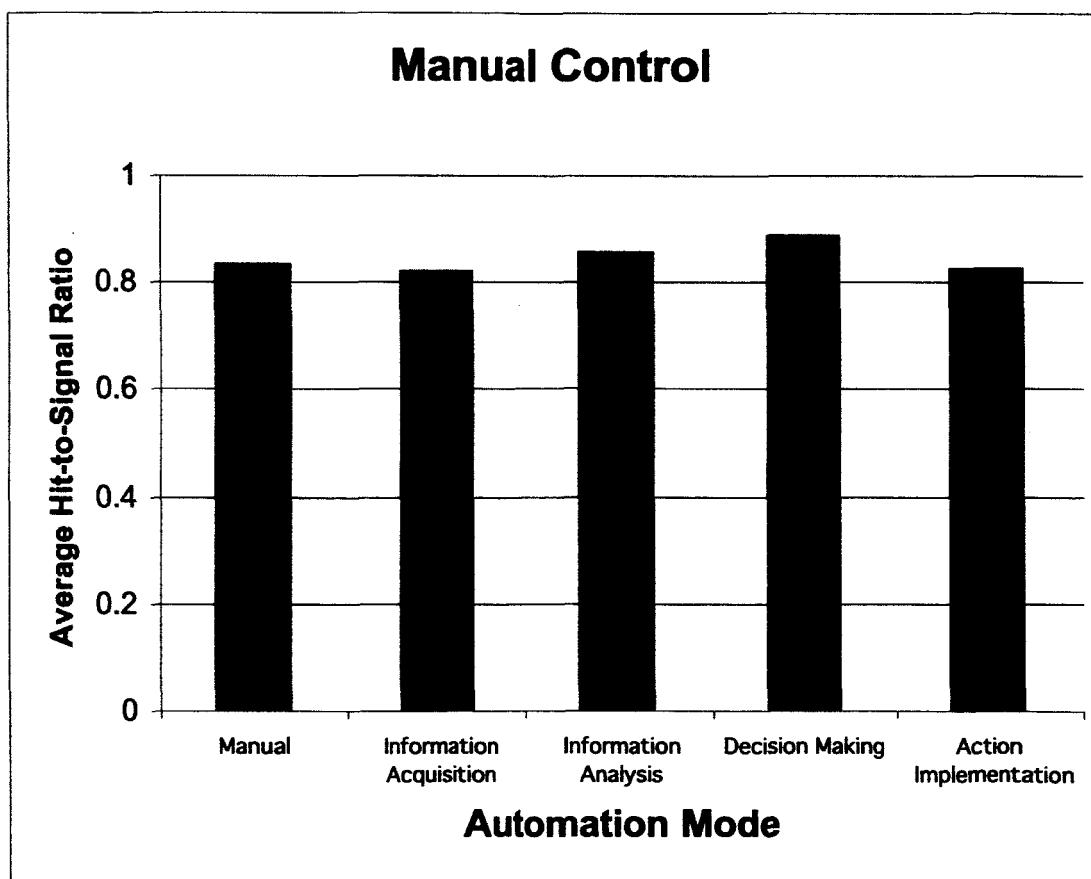


Figure 17. Secondary-task performance during manual control periods.

The pattern of secondary-task performance under the manual control periods was almost exactly opposite to that observed during the automated control periods. Duncan's test indicated that average workload was significantly lower under decision-making automation, as compared to workload during manual control periods in AA of the information acquisition and action implementation functions, as well as the completely manual control condition. It is possible that this is evidence of an automation carryover effect during these trials; when decision-making AA was applied and the recommendations for conflict avoidance were followed, the subsequent result was a lower workload when the simulation returned to manual control. This may have freed more attention for the secondary gauge task. This is in line with Laois & Giannacourou's (1995) finding that "significant workload reductions will be effected by aiding decision making and predictive activities more than by automation of routine data acquisition and communication activities".

Situation Awareness

Figure 18 summarizes the mean Level 1, Level 2, Level 3 and Total SA scores for both automated and manual control periods. A marginally significant effect of the mode of automation was found for Level 2 SA queries ($F(1,227)=3.51$, $p=0.0623$), indicating that subject comprehension was, on average, higher during manual control periods compared to automated control periods. This observation is in agreement with our hypothesis (E2H3), but the test statistic was not significant at the α -criterion we established for the research. In general, the observation supports the notion that introducing automation in ATC may remove the controller from the control loop (Endsley & Kaber, 1999) and lead to decrements in SA. Unlike in Experiment #1, there were no significant effects of mode of operation on the percent correct responses to Level 1 and 3 SA queries or for Total SA scores.

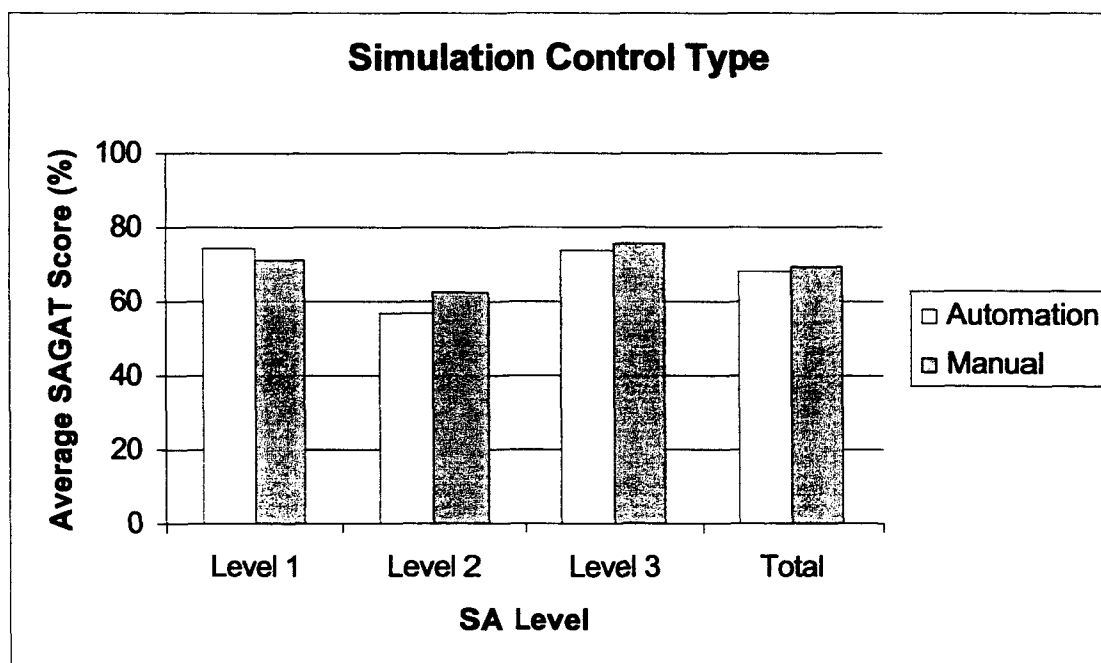


Figure 18. Mean SAGAT scores during manual and automation control periods

An ANOVA on the SA response measures revealed a significant effect of the specific forms of AA on Level 1 SA queries ($F(4,227)=3.78$, $p=0.0054$) and the Total SA score ($F(4,227)=2.7$, $p=0.0317$). This finding supports our expectation that the modified version of the SAGAT-based measurement technique was sensitive to the AA manipulations as part of the experiment. A significant Shapiro-Wilk's test ($p=0.013$) on the total SA score data suggested a potential violation of the normality assumption of the ANOVA. However, examination of a normal probability plot on the data did not reveal the potential departure from the normal distribution to be severe enough to merit transformation of the entire data set. No significant effects were found for Level 2 and Level 3 SA queries. Figures 19 and 20 show the average SAGAT scores under each mode of automation for Level 1 and Total SA respectively.

We also hypothesized (E2H3) that subjects would be better at responding to SA queries under information acquisition and manual control as compared to high-level automation. The TPA presented during information acquisition automation was expected to draw operator visual attention to the display and promote concentration. Duncan's test showed Level 1 SA to be significantly superior under information acquisition automation, compared to information analysis, decision making, action implementation, and manual trials ($p<0.05$). However, manual control was not found to increase Level 1 SA, as compared to the other automated conditions. The results on Level 1 SA also support hypothesis E2H4, which states that operator SA would be lower under decision making and information analysis automation.

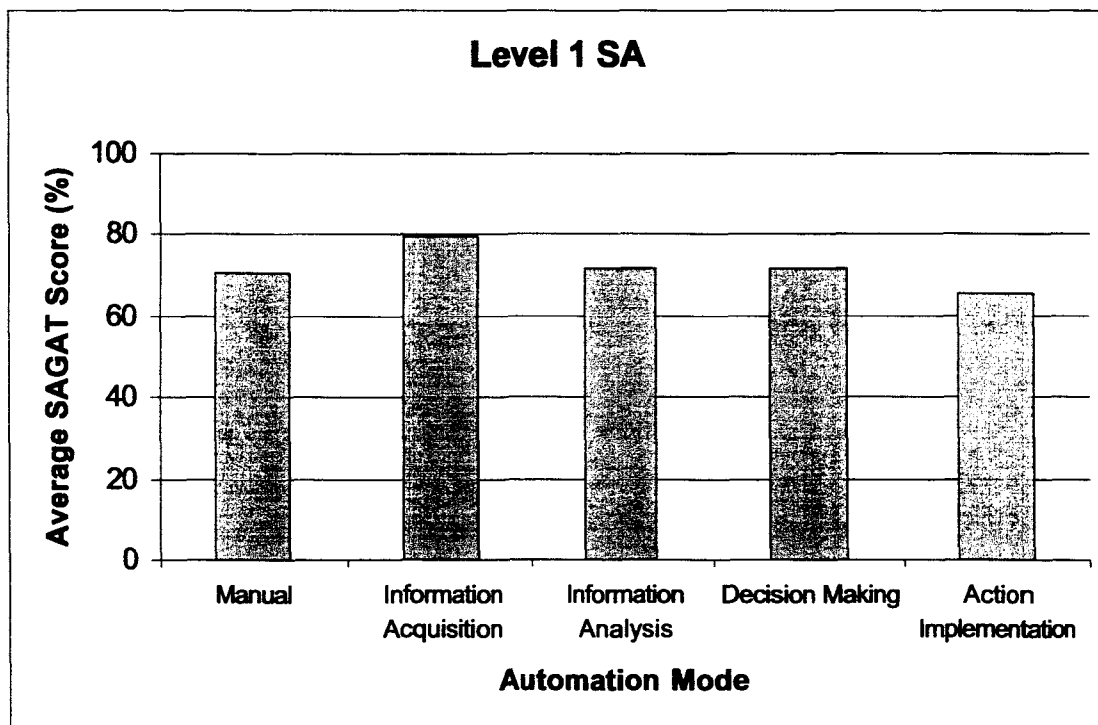


Figure 19. Mean Level 1 SAGAT scores for the different modes of automation

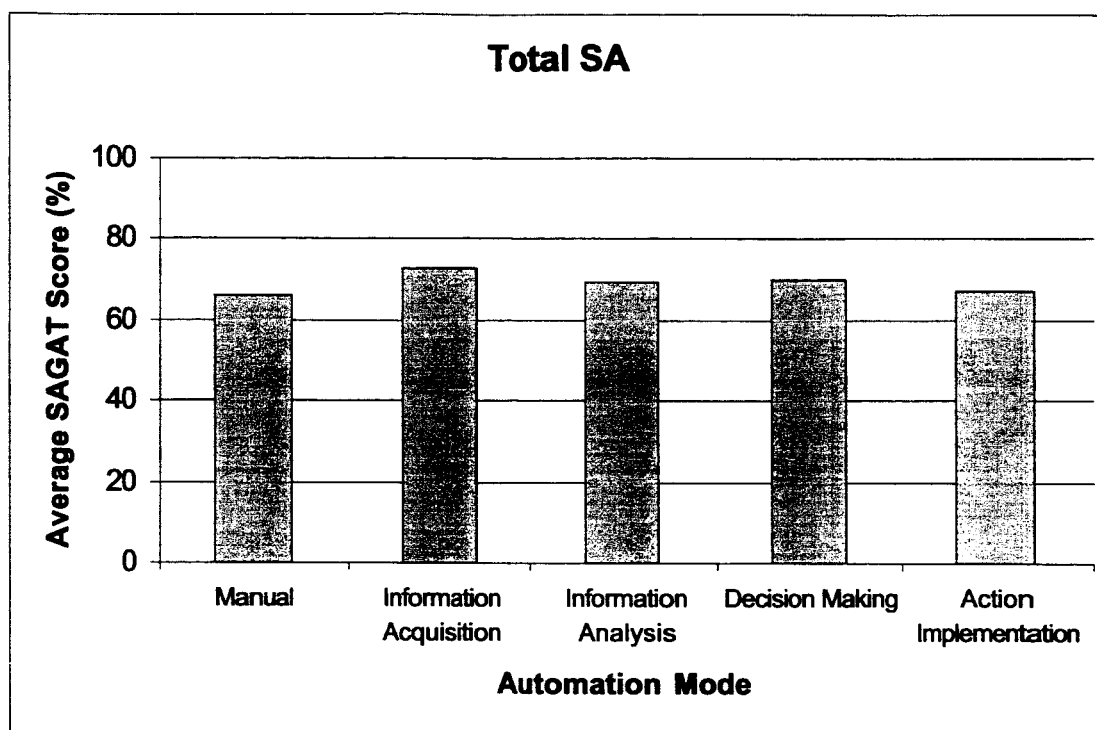


Figure 20. Mean total SAGAT scores for the different modes of automation

Duncan's test revealed information acquisition to be superior, in terms of Total SA, to action implementation automation ($p < 0.05$), which is also consistent with our hypothesis (E2H3). However, operator SA during manual trials was lower than SA during information acquisition trials, contrary to our previous finding that subject comprehension was better under manual control. In addition, SA during information analysis and decision making trials was not inferior to SA during other automation trials, counter to hypothesis E2H4.

In general, these results indicate that the modified SAGAT approach to measuring controller SA proved to be sensitive to differences among the AA conditions, particularly in terms of operator perceptual knowledge. By comparison with the SA measurement approach taken in Experiment #1, cueing subject recall of aircraft and using relevance weighting of aircraft, based on simulation events and recent controller actions, as part of the administration of SA queries, lead to observation of meaningful, significant differences in percent correct operator responses to Level 1 SA queries and in overall SA under the various automated ATC conditions.

Performance and Situation Awareness Correlation Analyses

A correlation analysis of the performance data revealed a significant positive correlation between conflicts and collisions ($r = 0.37148$, $p = 0.0007$). As one would expect, as the number of aircraft in conflict increased, the number of collisions also increased. Likewise, there was a significant negative correlation between the number of aircraft cleared and the number of collisions ($r = -0.25231$, $p = 0.024$). Following a collision, the aircraft that were conflicting were removed from the simulation display and new aircraft were generated at the periphery of the control sector.

Having to “start over” with new aircraft led to a decrease in the number of aircraft cleared during a trial with collisions.

A correlation analysis of the SA measures revealed a significant positive correlation between Level 2 and Level 3 SA ($r=0.23495$, $p=0.0359$). A significant positive correlation was also found between Level 1 SA and Total SA ($r=0.50687$, $p<0.0001$), Level 2 SA and Total SA ($r=0.61645$, $p<0.0001$), and Level 3 SA and Total SA ($r=0.72958$, $p<0.0001$). These findings indicate that a subject’s comprehension of the simulation was positively associated with the ability to project the future state of relevant aircraft. The results are in-line with Endsley’s (1995) theory on SA, specifically that higher levels of SA (comprehension and projection) are dependent upon perceptual knowledge and comprehension.

Of greater importance here were marginally significant positive correlations between Level 1 SA and cleared aircraft ($r=0.19466$, $p=0.0836$), Level 3 SA and cleared aircraft ($r=0.21388$, $p=0.0568$), and Total SA and cleared aircraft ($r=0.19754$, $p=0.079$). These findings reveal a positive association between subject perception and projection of the simulation and performance; that is, controller performance in the Multitask© simulation may be dependent upon their SA. Some other marginally significant correlations of SA and performance measures were observed but were not logical or explainable. The correlations of SA and performance are encouraging with respect to using a real-time measure of operator SA as a basis for triggering DFAs in adaptive systems. Our results suggest that SA in ATC maybe an alternative, yet converging measure of the state of the operator, and SA may be a viable trigger to AA, in lieu of using performance- or workload- matched AA.

5. Conclusions

Caveats

Simulation Fidelity

There are several limitations of this research that need to be noted and considered in interpretation of the findings, as well as the use of the results as a basis for systems design or applications development. First, Multitask© is a low-fidelity ATC simulation. Although its functions simulate cognitive requirements of real ATC, it has several limitations when compared to true ATC displays and functions. Multitask© is currently only capable of displaying aircraft in a 2-dimensional environment and the simulation does not require normal 3-dimensional operations (i.e., altitude knowledge, vertical separation, vertical approach procedures, etc.). The 3-dimensional property of ATC adds immensely to the task’s complexity and increases the amount of information needed by a controller in order to maintain aircraft separation. The vertical dimension may affect controller task performance, workload, and SA.

Secondly, the Multitask© simulation lacks certain ATC functions, such as verbal communications, flight strip management, emergency procedures, and dealing with pilot inadvertent deviations from prescribed routes. These functions within true ATC add immensely to task workload. In an attempt to simulate true ATC workload associated with these demanding

functions, the simulation speed in both experiments was increased to 2-times above real-time. However, it is possible that even this increase in simulation speed may not have raised task workload to the levels of real ATC operations.

Finally, the Multitask© simulation does not evoke in users the stresses associated with clearing aircraft that are carrying people. It is possible that this had a major impact on subject motivation and concentration in the simulation trials and, consequently, the workload, performance, and SA responses. In general, all the limitations of the simulation identified here subtract from the generalizability of the experimental results to the real ATC domain.

Subject Training

Based on the unique characteristics of the experimental task, novice operators were used in this study. Although all subjects completed a thorough training program and they may have been experts at performing the simulation, they still did not represent certified FAA air traffic controllers. Although this study revealed significant findings on AA in the ATC simulation and the implications for controller SA, the results may not be generalizable to expert ATC controllers due to the use of a novice subject population.

SA Measurement Technique Problems

Finally, as previously mentioned, the first experiment employed a SA measure without relevance weighting of aircraft, based on the events of the simulation scenario. It is possible that this approach to SAGAT did not account for the varying emphasis that controllers may place on aircraft, compromising the validity of the measurement technique for the ATC domain. The results of the second experiment provided additional evidence to support this notion. Our modified version of SAGAT, involving cued recall of aircraft, and aircraft relevance weighting, proved to be sensitive to differences among the AA conditions in terms of SA.

Design Implications for Future ATC Automation

Contributions to AA Research

The results of this study may serve as an applicable guide for the design of future ATC automation, and aviation systems automation, in general. We completed two experiments with results demonstrating differential effects of AA applied to various information processing functions as part of ATC. We also observed that control performance varies among automated and manual control periods as part of AA. With respect to automated control, our results across experiments were consistent, indicating that higher levels of automation, involving information analysis assistance (identification of potential aircraft conflicts), may be detrimental to successfully clearing aircraft. However, as one might expect, the same form of automation appears to improve conflict detection. Although the automation made conflicts salient to controllers, the fact that they were removed from the control loop in terms of closely monitoring aircraft flight parameters and projecting the conflicts themselves seems to have a negative implication on the issuance of appropriate clearances to land aircraft. Beyond this, we observed that lower levels of automation, including action implementation and information acquisition,

which support controllers in terms of clearing aircraft and gathering information on flight parameters, promote performance, specifically, the number of aircraft clearances. With respect to manual control periods as part of AA of various ATC functions, we also observed consistent results across experiments demonstrating that, when automation is applied to the information analysis function of the task, in subsequent manual control periods controllers appear to do better at successfully clearing aircraft than when using lower level automation providing for automatic clearances.

With respect to the approach to AA that we explored through this research, we found the secondary gauge monitoring task to be a sensitive indicator of workload fluctuations in the primary ATC simulation, as has been observed in other studies (Kaber, Prinzel et. al., 2002; Kaber & Clamann, 2003). In general, we observed that when AA is applied to the higher-order information processing functions as part of ATC, secondary task performance decreases, or workload increases. We have previously made the inference that the complexity of these forms of automation may be such that they draw additional operator visual attention, or the recommendations being made by the decision aids require time for the operators to evaluate and to make comparison with their own projections of future clearances. Consequently, there is less time for them to attend to the secondary gauge task, indicating an increase in workload in the primary task. The lower-order forms of automation, including information acquisition and action implementation, consistently lead to increases in secondary task performance or reductions in workload. This is most likely because these modes of automation either provide operators with additional data on aircraft (in the same area as the radarscope) or they reduce the number of interface actions required of controllers in order to successfully land an aircraft. Interestingly, the pattern of findings on secondary task performance during manual control periods of the primary task, as part of AA trials, is almost exactly opposite to the pattern of findings observed during the automated control periods. Controller exposure to lower levels of automation appears to lead to higher workload for subjects during the manual control periods, and secondary task performance is worse during periods of high-level automated control of ATC functions. It is possible that as a result of subjects utilizing advanced automation, such as decision aiding and following computer recommendations for conflict resolution, in advance of manual control periods ultimately resulted in lower workload when the simulation did shift to fully manual control. In general, these findings demonstrate that a secondary task can be an effective trigger for DFAs in the context of ATC. Other forms of secondary tasks that are more closely related to actual ATC functions may also be useful for this purpose. (We say more about this in the future research section.)

Contributions to SA Research

With respect to SA, this research has advanced the state of measurement for the purpose of describing the implications of AA on complex task performance. It has also provided detailed information on fluctuations in operator perceptual knowledge, comprehension, and projection when AA is applied to ATC information processing functions including information acquisition, information analysis, decision making and action implementation. Both experiments we conducted as part of this project revealed some effect of the general mode of control (i.e., manual versus automated) on controller SA. However, this occurred at different levels of SA, specifically subject projection abilities were impacted in Experiment #1 and comprehension

abilities were impacted in Experiment #2. However, the investigation of the effect of specific forms of AA on the SA response measures, as part of Experiment #1, revealed no significant effects. Furthermore, when we analyzed the impact of the various forms of AA on the percent correct responses to specific SA queries, we also did not find any sensitivity of the measurement approach.

Fortunately, our revision to the SAGAT-based approach to measuring SA as part of the second experiment proved to be effective in terms of assessing the impact of specific forms of AA on ATC performance. Using queued recall of aircraft, and establishing relevance weights for various aircraft at the time of SAGAT freezes, appeared to cause the SA response measures, including Level 1, 2, and 3 SA, to be sensitive to the AA manipulations. As we discussed, this measurement approach is similar to that explored by Hauss and Eyferth (2003), which also showed sensitivity to experimental manipulations in the context of an ATC task, relative to a conventional SAGAT-based approach. Specifically, our findings on Level 1 SA revealed subject perception to be significantly superior under low levels of automation, including information acquisition, which provided subjects with additional information on aircraft flight parameters. We did hypothesize that manual control would produce SA superior to the automated conditions; however, this was not the case. The finding may be attributed to high levels of subject workload when manually issuing clearances versus having any automated assistance whatsoever.

To summarize, the experiments we presented here resulted in a new approach to measuring SA that is sensitive to AA manipulations in the context of ATC. This was one of the primary objectives of the work. Beyond this, the measure was used to describe in detail the differential effects of AA applied to the various information processing functions, as part of ATC, on controller perception, comprehension and projection.

Future Research

Automation is currently implemented in ATC to alleviate workload placed on controllers while improving ATC performance and efficiency. Likewise, some manual control remains necessary to prevent controller OOTLUF and the associated negative human performance consequences (loss of SA, skill decay). Adaptive automation has been proposed as an alternative to conventional automation that may provide the benefits of moderating operator workload and consideration of OOTL performance problems. However, the precise relationship between the duration of DFAs and the number of automation or manual control periods needed within a task are still unknown. Further research using various secondary-task criteria as bases for DFAs as part of AA is needed to describe the relationships between various proportions of manual and automated control to primary task performance and SA. In addition to defining these relationships, an optimum amount of automated and manual control time during AA should be found to maximize air traffic controller performance and SA within a robust ATC simulation.

The second experiment as part of this research validated a modified SAGAT-based measure of SA, incorporating relevance weighting of aircraft based on simulation events and recent controller actions, for air traffic control scenarios. Our approach was similar to Hauss and Eyferth's (2003) SA measurement technique; however, they used expert controller evaluations of scenario replays in order to identify aircraft that "should've" been most important to controllers

at the time of simulation freezes. In our study, expert experimenters used a form of a CDA and CAA, as well as their own observations on recent clearances issued by subjects, to determine in real-time, which aircraft were to be used as bases for SA queries during freezes. A hierarchy of events and actions, including recently detected conflicts and recent or expected clearance changes, was used to establish the relevance weighting of aircraft. This method appeared to be sensitive to detecting differences in controller SA due to DFAs as part of AA of various ATC information processing functions and to reliably reveal changes in SA.

Hauss and Eyferth (2003) attribute the need for a weighted SA measure to the unique complexity of the ATC environment requiring controllers to manage a large amount of information using an event based mental representation. As a result, varying levels of relevance are applied to each aircraft within the controller's airspace. It is possible that relevance of stimuli or events to human performance in other domains, such as driving and piloting, may be important to consider in developing SA measures for studying, for example, in-vehicle highway systems or cockpit automation effects of pilot SA. Research is needed to identify other environments in which weighted SA measures may be applicable and useful. Subsequently, these SA measures may be used to assess the effects of AA of various human information processing functions on SA within different contexts.

Once the effects of AA on SA are fully understood, future research is needed to develop a real-time probe measure of SA, which may serve as a basis for triggering DFAs in complex systems control. Although this was one of the initial objectives of this research, the lack of sensitivity of the prototype SA measure, observed during the first experiment, made the objective unachievable within the performance period of the project.

In the context of an ATC simulation, like Multitask©, it is possible to introduce communication errors between virtual aircraft pilots and subject controllers and to use the errors to assess controller SA. For example, an aircraft may be non-responsive to an ATC flight query or command of speed change, holding pattern, re-route, etc. Whether a controller observes this, or reacts to the pilot error, can be considered an indicator of controller alertness of simulated events or awareness of aircraft states. Another type of simulation event allowing experimenters to probe controller awareness might be an aircraft responding erroneously to a controller clearance command. For example, confirming a speed change when cleared for a holding pattern or vice versa. Aircraft might also confirm incorrect speed changes, holding patterns, etc., in response to a controller command. Again, observing whether a controller reacts to such errors, how quickly they respond, and the effectiveness of their action, can all be considered probes of SA. Other types of ATC simulation errors that would allow for probing of controller SA include aircraft providing unsolicited information to controllers, such as confirmation or refusal of clearances that have not been issued. Unfortunately, these types of communication errors are not uncommon in real-world ATC because of controller and aircraft pilot workload. Observing if a subject controller detects the unsolicited communication from an aircraft, and how they address it, may provide insight into their perception and comprehension of aircraft flight parameters. For example, when using the Multitask© simulation, depending upon whether probe events are detected in the communication history display, it should be possible to determine the level of operator perceptual knowledge (Level 1 SA). In order to assess operator comprehension of aircraft states, it will be necessary to record their actions at the ATC simulation interface and

relate them to the specific probes presented during the simulation. One method of assessing operator ability to project future simulation states might be to present probe events representing errors of commission by aircraft accepting clearances that a controller should have issued.

This would be one approach to delivery a real-time probe measure of SA in ATC. All of the probes described above could appear to be imbedded in the primary task of simulation operators and may serve as unobtrusive measures of controller SA. However, there might be implications for subject workload and, consequently, an influence on the pattern of AA, if workload is used as a basis for triggering DFAs.

As previously speculated, using an SA-based trigger of function allocations between an operator and automation may produce superior results compared to, for example, a workload based trigger of DFAs. The criterion for this type of AA design would entail allocating control to automation only when the controller is fully "in-the-loop" and has achieved sufficient SA. However, if the controller loses "the picture" (SA), manual control could be reinstated in order to improve perception, comprehension, and projection of system status.

It would be necessary to empirically assess the effectiveness of such an approach to AA of ATC information processing functions, as other studies (Kaber, Prinzel et al., 2002; Clamann & Kaber, 2003) have demonstrated performance benefits of AA applied to lower-order sensory/response functions, but Kaber & Endsley (2004) showed decrements in operator SA when exposed to the same forms of automation. Beyond this, at least one study has demonstrated improvements in operator SA in a dynamic control task due to intermediate levels of automation applying computer assistance to the planning and implementation aspects of the task (Kaber & Endsley, 2004). To conduct an experiment using, for example, real-time probes as basis for assessing SA and triggering DFAs, it would be necessary to establish criterion levels of SA (average levels of operator perceptual knowledge, comprehension of system states and ability to project system states) for the various modes of ATC automation to be tested. During the experiment, the DFAs as part of AA conditions would be facilitated based on real-time analysis of operator responses to probe events involving simulated aircraft, and comparison of calculated operator SA with the established SA criteria. If operator SA during test trials substantially deviated from the SA criteria, automated or manual control allocations would be invoked. For example, if a percent decrease in operator SA from one probe event to the next, during Multitask© performance, lead to overall operator SA to fall below a lower criterion indicative of "poor" comprehension, we might consider this to mean operator OOTLUF, and manual control allocations to the operator would occur. If operator SA met or exceeded the upper SA criterion level for a specific automation condition, automated control allocations could be provided to relieve the operator of some task workload.

In general, this represents a potential SA-based approach to an AA triggering mechanism that needs to be explored in future research. Since a SA-based trigger would capture operator cognitive state and then relate that state to DFAs, it could be expected that using probe events may improve the effectiveness of AA as applied to higher-order cognitive processes, as compared to using a performance- or workload-based approach. The previous research as part of this program has shown that a workload-based approach to AA (Kaber, Prinzel et al., 2002; Kaber & Clamann, 2003) may promote the effectiveness of AA applied to lower-order

information processing functions, such as information acquisition and action implementation. This may be due to the fact that the secondary-task measure of workload is not sensitive to changes in higher-level cognition. It may also be due to the fact that performance measures do not provide direct insight into cognitive load. However, the probe events proposed for use here may directly tap various aspects of operator SA based on their design and promote the effectiveness of AA applied to cognitive tasks.

References

- Bolstad, C., & Endsley, M. R. (2000). Shared displays and team performance. In D. B. Kaber and M. R. Endsley (eds.), *Human Performance, Situation Awareness and Automation: User-centered Design for the New Millennium* (pp. 311-317). Madison, WI: Omnipress.
- Clamann, M. P. and Kaber, D. B. (2003). Authority in adaptive automation applied to various stages of human-machine system information processing. In *Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 543-547). Santa Monica, CA: Human Factors and Ergonomics Society.
- Clamann, M. P., Wright, M. C. and Kaber, D. B. (2002). Comparison of performance effects of adaptive automation applied to various stages of human-machine system information processing. In *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 342-346). Santa Monica, CA: Human Factors and Ergonomics Society.
- Dillingham, G. L. (1998). *Air traffic control: Evolution and status of FAA's automation program*. (Tech. Report GAO/T-RCED/AIMD-98-85). Washington, DC: United States General Accounting Office.
- Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32-64.
- Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors*, 37(1), 65-84.
- Endsley, M. R. (1995c). Towards a new paradigm for automation: Designing for situation awareness. In: *Proceedings of the 6th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems*, pp. 421-426. Cambridge, MA: MIT.
- Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum.
- Endsley, M. R., & Jones, D. G. (1995). *Situation awareness requirements analysis for TRACON air traffic control* (Tech. Report TTU-IE-95-01). Atlantic City, NJ: Technology Center, Federal Aviation Administration.
- Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.
- Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37, 381-394.
- Endsley, M. R. & Rodgers, M. D. (1994). *Situation awareness information requirements for en route air traffic control*. (Tech. Report DOT/FAA/AM-94/27). Washington, DC: Office of Aviation Medicine, United States Department of Transportation, Federal Aviation Administration.
- Gronlund, S. D., Ohrt, D. D., Dougherty, M. R. P., Perry, J. L., & Manning, C. A. (1998). *Aircraft importance and its relevance to situation awareness*. (Tech. Report DOT/FAA/AM-98/16). Springfield, VI: National Technical Information Service.
- Hauss, Y. & Eyferth, K. (2003). Securing future ATM-concepts' safety by measuring situation awareness in ATC. *Aerospace Science and Technology*, 7, 417-427.
- Hilburn, B. G., Jorna, P. Byrne, E. A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua & J. M.

- Koonce (Eds.), *Human-automation interaction: Research and practice* (pp. 84-91). Mahwah, NJ: Lawrence Erlbaum Assoc., Inc.
- Hilburn, B., Molloy, R., Wong, D., & Parasuraman, R. (1993). Operator versus computer control of adaptive automation. In R. S. Jensen & D. Neumeister (eds.), *Proceedings of the 7th International Symposium on Aviation Psychology* (pp. 161-166). Columbus, OH: Department of Aviation, The Ohio State University.
- Hopkin, V. D. (1998). The impact of automation on air traffic control specialists. In M. W. Smolensky & E. S. Stein (Eds.), *Human factors in air traffic control* (pp. 391-419). San Diego, CA: Academic Press.
- Jones, D.G. (2000). Subjective Measures of Situation Awareness. In M.R. Endsley & D.J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 113-128). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Jones, D. G. & Endsley, M.R. (2000). Can real-time probes provide a valid measure of situation awareness? In D. B. Kaber & M. R. Endsley (Eds.), *Human Performance, Situation Awareness and Automation: User-centered Design for the New Millennium* (245-250). Atlanta, GA: SA Technologies.
- Kaber, D. B. and Clamann, M. P. (March 2003). *Authority in adaptive automation applied to various stages of human-machine system information processing*. (Final Rep.: NASA Langley Research Center Grant #NAG-1-02056). Hampton, VA: NASA Langley.
- Kaber, D. B. & Endsley, M. R. (1997). The combined effect of level of automation and adaptive automation on human performance with complex, dynamic control systems. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 205-209). Santa Monica, CA: Human Factors and Ergonomics Society.
- Kaber, D. B. and Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 5(2), 113-153.
- Kaber, D. B., Onal, E., & Endsley, M. R. (1999). Level of automation effects on telerobot performance and human operator situation awareness and subjective workload. In M. Scerbo & M. Mouloua (Eds.), *Automation technology and human performance* (pp. 165-170). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kaber, D. B., Prinzel, L. J., Wright, M. C. & Clamann, M. P. (2002). *Workload-matched adaptive automation support of air traffic controller information processing stages* (Tech. Pub.: NASA/TP-2002-211932). Washington, DC: NASA.
- Kaber, D. B., & Riley, J. (1999). Adaptive automation of a dynamic control task based on secondary task workload measurement. *International Journal of Cognitive Ergonomics*, 3(3), 169-187.
- Kaber, D. B., Riley, J. M., Tan, K., & Endsley, M. R. (2001). On the Design of Adaptive Automation for Complex Systems. *International Journal of Cognitive Ergonomics*, 5(1), 37-57.
- Kaber, D. B., Wright, M. C. and Hughes, L. E. (February 2002). *Automation-state changes and sensory cueing in complex systems control* (Final Rep.: Office of Naval Research Award #N0001401010402). Arlington, VA: Office of Naval Research.
- Laois, L. & Giannacourou, M. (1995). Perceived effects of advanced ATC functions on human activities: Results of a survey on controllers and experts. In *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 392-397). Columbus, OH: The Ohio State University.

- Leroux, M. (1993). The role of expert systems in future cooperative tools for air traffic controllers. In *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 335-340). Columbus, OH: The Ohio State University.
- Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our Capacity for processing information. *The Psychological Review*, 63, 81-97.
- Montgomery, D. C. (2001). *Design and Analysis of Experiments*. New York: John Wiley & Sons, Inc.
- National Research Council (1997). *Flight to the future: Human factors in air traffic control*. Washington, DC: National Academy Press.
- National Research Council (1998). *The future of air traffic control: Human operators and automation*. Washington, DC: National Academy Press.
- Neter, J., Wasserman, W., & Kutner, M. H. (1990). *Applied Linear Statistical Models*. Boston, MA: Irwin, Inc.
- Niessen, C., Eyferth, K., & Bierwagen, T. (1999). Modeling cognitive processes of experienced air traffic controllers. *Ergonomics*, 42(11), 1507-1520.
- Nolan, M. S. (1999). *Fundamentals of air traffic control* (3rd ed.). Pacific Grove, CA: Brooks/Cole.
- Nunes, A. (2003). The impact of automation use on the mental model: findings from the air traffic control domain. In *Proceedings of the Human Factors & Ergonomics Society 47th Annual Meeting* (pp. 66-70). Santa Monica, CA: Human Factors and Ergonomics Society.
- Parasuraman, R. (2000). Designing automation for human use: empirical studies and quantitative models. *Ergonomics*, 43(7), 931-951.
- Parasuraman, R. & Byrne, E. A. (2002). *Automation and human performance in aviation*. In P. Tsang & M. Vidulich (Eds.), *Principles and practices of aviation psychology* (pp. 311-356). Mahwah, NJ: Lawrence Erlbaum.
- Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1993). Adaptive function allocation reduces performance costs of static automation. In R. S. Jensen & D. Neumeister (eds.), *Proceedings of the 7th International Symposium on Aviation Psychology* (178-185). Columbus, OH: Department of Aviation, The Ohio State University.
- Parasuraman, R., & Riley, V. A. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39, 230-253.
- Parasuraman, R., Wickens, C. D., & Sheridan, T. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics*, 30(3), 286-297.
- SAS Institute Inc. (1990). *SAS/STAT® User's Guide* (Version 6, 4th ed). Cary, NC: SAS Institute, Inc.
- Taylor, R. M. (1990). *Situational awareness rating technique (SART): The development of a tool for aircrew systems design*. *Situational Awareness in Aerospace Operations* (AGARD-CP-478 ed., pp. 3/1-3/17). Neuilly Sur Sience: NATO-AGARD.
- Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York, NY: HarperCollins Inc.
- Wickens, C. D. (2000). *Imperfect and unreliable automation and its implications for attention allocation, information access, and situational awareness* (Tech. Report ARL-00-10/NASA-00-2). Savoy, IL: Aviation Research Lab, Institute of Aviation, University of Illinois.
- Wickens, C.D., & Hollands, J.G. (2000). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River: Prentice Hall Inc.

Appendix A: GDTA for Multitask© simulation.

A GDTA is conducted by first identifying the major goals for a task. Secondly, subgoals are determined, which are essential for accomplishing the primary goal. Subsequently, the critical decisions associated with each subgoal are identified and used as a basis for establishing operator SA requirements to complete the task (Endsley & Jones, 1995). These requirements focus not only on what data the controller needs, but also on how that information is integrated or combined to address each decision (Endsley & Rodgers, 1994, p. 4).

Endsley and Rodgers (1994) identify four caveats that relate to a GDTA:

- (1) At any given time, more than one goal or subgoal may be operating, although these will not always have the same priority. The analysis does not assume any prioritization among goals, or that each subgoal within a goal will always be relevant.
- (2) The analysis is based on goals or objectives, and is as technology-free as possible. How the information is acquired is not addressed.
- (3) The analysis seeks to determine what controllers would ideally like to know in order to meet each goal.
- (4) Static knowledge, such as procedures or rules for performing tasks, is outside the bounds of this analysis.

(Endsley & Rodgers, 1994, p. 4-5)

The following GDTA describes the goals and information requirements required to successfully clear and land aircraft at one of two airports in the Multitask© simulation. The GDTA was prepared using the methods described by Endsley and Rodgers (1994) and Endsley and Jones (1995). This analysis assumes expert performance (i.e., operator errors are not considered). Each major decision, and the SA requirements associated with a subgoal, represents particular levels of information processing (perception, comprehension, or projection). (The SA requirements were used as a basis for developing SAGAT queries (see Appendix B).)

At the end of the GDTA, there are plans defined for operators to accomplish the subgoals as part of the ATC simulation. The plans are representative of an aspect of traditional task analysis. Finally, a detailed description of the Multitask© goals and subgoals is provided for reference in reviewing the GDTA.

<u>Goal</u>	<u>Subgoal</u>	<u>Decisions/SA requirements</u>	<u>Level of SA</u>
-------------	----------------	----------------------------------	--------------------

(Note: Some decisions or questions make reference to other information/SA requirements. These requirements are emboldened.)

0. land aircraft safely

1. acquire aircraft info

1.1 locate aircraft

what display sector is the aircraft located?	Level 1
aircraft position	
how many other aircraft are located in that sector?	Level 1
aircraft position	
location of other aircraft	

1.2 contact aircraft

1.3 verify

is the aircraft queried (contacted)?	Level 1
aircraft icon (flashing or solid)	
aircraft parameters	

1.4 acquire information

what is the aircraft type?	Level 1
aircraft call sign	
aircraft speed	
apparent velocity of aircraft	
what is the aircraft call sign?	Level 1
aircraft call sign	
what is the aircraft speed?	Level 1
aircraft speed	
apparent velocity of aircraft	
what is the destination airport?	Level 1
destination airport	
apparent aircraft route	
what is the destination runway?	Level 1
destination runway	
what is the aircraft heading?	Level 2
destination airport	
apparent aircraft route	
what is the distance to destination?	Level 2
destination airport	

aircraft position (1.1)	
what is the distance to nearest holding fix?	Level 2
aircraft position (1.1)	
location of nearest holding fix	
what is the heading to nearest holding fix?	Level 2
aircraft position (1.1)	
location of nearest holding fix	
is aircraft speed representative of type?	Level 2
aircraft call sign	
aircraft speed	
apparent velocity of aircraft	

2. identify potential conflicts

2.1 determine current relationships between aircraft

what is the distance to other aircraft?	Level 2
aircraft position (1.1)	
location of other aircraft (1.1)	
what is the heading to other aircraft?	Level 2
aircraft position (1.1)	
location of other aircraft (1.1)	
do aircraft meet or exceed lateral separation?	Level 2
aircraft position (1.1)	
location of other aircraft (1.1)	
aircraft destination runway (1.4)	
other aircraft destination runways (1.4)	

2.2 predict future information

what sector will aircraft be located in the future?	Level 3
aircraft position (1.1)	
aircraft destination (1.4)	
aircraft speed (1.4)	
apparent aircraft route	
will aircraft overtake each other?	Level 3
aircraft position (1.1)	
aircraft destination (1.4)	
aircraft speed (1.4)	
apparent aircraft route	
other aircraft positions (1.1)	
other aircraft destinations (1.4)	
other aircraft speeds (1.4)	
other apparent aircraft routes	
do aircraft paths cross each other?	Level 3

aircraft position (1.1) aircraft destination (1.4) aircraft speed (1.4) apparent aircraft route other aircraft positions (1.1) other aircraft destinations (1.4) other aircraft speeds (1.4) other apparent aircraft routes	
when will aircraft converge?	Level 3
aircraft position (1.1) aircraft destination (1.4) aircraft speed (1.4) apparent aircraft route other aircraft positions (1.1) other aircraft destinations (1.4) other aircraft speeds (1.4) other apparent aircraft routes	
where will aircraft converge?	Level 3
aircraft position (1.1) aircraft destination (1.4) aircraft speed (1.4) apparent aircraft route other aircraft positions (1.1) other aircraft destinations (1.4) other aircraft speeds (1.4) other apparent aircraft routes	
when will aircraft arrive at airport?	Level 3
aircraft position (1.1) destination airport (1.4) aircraft speed (1.4) apparent aircraft velocity	

3. decide which aircraft clearance needed

3.1 choose aircraft to manipulate

which aircraft will need to be manipulated?	Level 3
potential conflict (2.2) aircraft capabilities projected effect on other aircraft	

3.2 choose operation

which operation should be used?	Level 3
potential conflict (2.2)	

operations, which can be used
aircraft capabilities
projected effect on other aircraft

4. provide clearance to chosen aircraft

4.1 select aircraft

4.2 select clearance

4.3 submit

4.4 verify

is aircraft conforming to assigned parameter(s)?	Level 1
changes to aircraft parameters (speed, route, etc)	
is aircraft conforming fast enough?	Level 1
aircraft position (1.1)	
aircraft destination (1.4)	
aircraft speed (1.4)	
apparent aircraft route	
other aircraft positions (1.1)	
other aircraft destinations (1.4)	
other aircraft speeds (1.4)	
other apparent aircraft routes	
are other aircraft diverting without clearance?	Level 1
changes to aircraft parameters (speed, route, etc)	
reason for non-conformance?	Level 2
aircraft position (1.1)	
operations, which can be used	
is aircraft conflict resolved?	Level 2
aircraft position (1.1)	
aircraft destination (1.4)	
aircraft speed (1.4)	
apparent aircraft route	
other aircraft positions (1.1)	
other aircraft destinations (1.4)	
other aircraft speeds (1.4)	
other apparent aircraft routes	
is aircraft resuming?	Level 1
changes to aircraft parameters (speed, route, etc)	

Plan 1: do 1.1 – 1.2 – 1.3 – 1.4.

if aircraft is still flashing, then repeat.

Plan 2: based on 2.1 and 2.2, do two aircraft appear to be converging?
if yes, then perform Plan 3.
if no, repeat Plan 1.

Plan 3: do 3.1 – 3.2, then perform Plan 4.

Plan 4: do 4.1 – 4.2 – 4.3 – 4.4.
if aircraft is non-compliant, then repeat Plan 2.
If aircraft is compliant and operation was 'hold', repeat Plan 2 until there is no conflict, then issue 'resume' clearance.

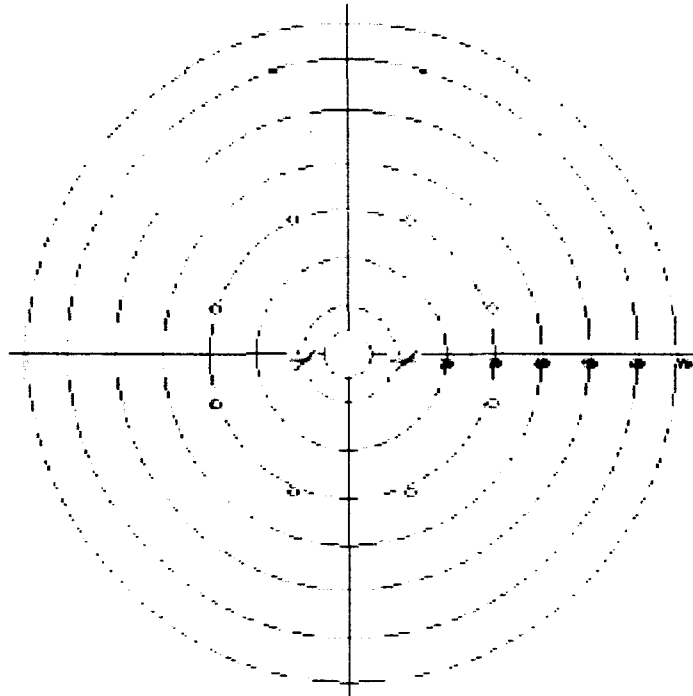
Descriptions of ATC goals in the context of the Multitask© simulation:

0. Land Aircraft Safely – The overarching goal of Multitask© is to land the aircraft safely at either of the two airports. Actions that constitute unsafe flight are: (1) potential aircraft conflicts, and (2) actual aircraft collisions. The task goal is accomplished by successfully contacting the aircraft, analyzing each aircraft's flight parameters, and, if needed, issuing additional aircraft clearances.
1. Contact aircraft – Locate and maintain contact with aircraft on the display.
 - 1.1. Locate aircraft – Find aircraft on the display and place cursor on aircraft.
 - 1.2. Contact aircraft – Double-click left mouse button on top of aircraft. Click on 'Query' in the control box. Wait for 'Processing...' to finish.
 - 1.3. Verify – Confirm that the aircraft icon is no longer flashing. If the icon is still flashing, then repeat Steps 1.1 and 1.2.
 - 1.4. Acquire aircraft information:
 - Aircraft type – Military, commercial, or private.
 - Aircraft call sign – Alpha-numeric designation.
 - Speed – Dependent on aircraft type.
 - Airport – A1 or A2.
 - Runway – R1 or R2 at the designated airport.
 - Location – In relation to airports and other aircraft.
 - Heading – Based on icon movement.
2. Identify potential conflicts – Operators are required to determine if a potential conflict with another aircraft is likely.
 - 2.1. Determine relationships between aircraft and two possible destinations.
 - 2.2. Predict future flight parameters of aircraft and determine if a change in aircraft parameters is required.
3. Decide which aircraft clearance(s) is/are needed – If a conflict is determined likely, the operator must decide the correct actions to take.
 - 3.1. Identify aircraft – Which aircraft need(s) to be cleared?
 - 3.2. Identify operation – Which clearance should be used to resolve conflict?
 - Reduce speed
 - Hold
 - Change airport
 - Change runway
4. Issue clearance – The proper operations are performed to resolve conflict.
 - 4.1. Select aircraft – Double-click left mouse button on top of appropriate aircraft.
 - 4.2. Select clearance – Chose the correct operation from the control box.
 - 4.3. Submit – Press 'Submit' on the control box.
 - 4.4. Verify clearance – Verify that the appropriate aircraft conformed to the assigned parameters (both from the control box script and the visual display).
 - 4.5. If the aircraft was issued a 'hold' clearance, a 'resume' clearance is required after the conflict is resolved.

Appendix B: SAGAT queries.

Subjects marked on a graphic of the Multitask© radarscope (see below) where each aircraft was located at the time of a simulation freeze. Specifically, they wrote the number 1-7 on the scope at the positions of the aircraft. Subject answers to follow-up Level 1, 2 and 3 SA queries were based on the numbers they assigned to each aircraft on the graphic. Subjects could refer back to this diagram while answering queries. The specific instructions to subjects for completing the graphic were as follows:

Using the following diagram, indicate where the 7 aircraft are currently located by randomly assigning a number 1-7 to each aircraft and writing the number on the graphic.



The table below presents all the queries posed to subjects during SAGAT freezes. Subjects responded to each query for every aircraft they recalled as being present on the display at the time of the freeze. There was only one correct answer for each aircraft based on the task situation at the time of the freeze. If necessary, response criteria were also identified for subjects in conjunction with presentation of the query. During each freeze a random selection of 6 of the following 18 questions was presented to a subject.

Complete list of SAGAT queries for experiments.

Level 1 SA	A. What is the aircraft's call sign?
	B. Is the aircraft flying at its original or reduced speed?
	C. What is the aircraft's destination airport? (A1 or A2)
	D. What is the aircraft's destination runway? (R1 or R2)
	E. What is the aircraft's type? (Military, Commercial, or Private)

	F. Has the aircraft already been queried/contacted? (Y or N)
Level 2 SA	A. What is the aircraft's heading (in degrees)?* (Criterion: Subject's answer must be within 20 deg. of actual heading for it to be graded as correct.)
	B. What is the distance of the aircraft from its destination (in nm)? (Criterion: Subject's answer must be within 5 nm of actual position for it to be graded as correct.)
	C. What is the aircraft's distance to the nearest aircraft (in nm)? (Criterion: Subject's answer must be within 5 nm of actual position for it to be graded as correct.)
	D. What is the aircraft's heading to the nearest aircraft (in degrees)?* (Criterion: Subject's answer must be within 20 deg. of actual heading for it to be graded as correct.)
	E. Does the aircraft meet or exceed lateral separation requirements? (Y or N)
	F. The aircraft shares its assigned route with how many other aircraft? (0-7)
Level 3 SA	A. What clearance(s) will be required? (Hold, Reduce Speed, Change Airport, Change Runway, Resume, None)
	B. If a clearance change is needed, when will the change be issued (in min. from now)? (Criterion: Subject's answer must be within 2 min. of actual time for it to be graded as correct.)
	C. Without a clearance change, will the aircraft overtake any other aircraft? (Y or N)
	D. Without a clearance change, will the aircraft's path cross another aircraft's path? (Y or N)
	E. When will the aircraft arrive at the destination airport (in min. from now)? (Criterion: Subject's answer must be within 2 min. of actual time for it to be graded as correct.)
	F. Rank the aircraft in the order that they will land/sequence. (1-7)

* - The following diagram was presented when subjects answered Level 2 SA, Items A and D.

